Memories in Computers
Part 2: DDR SDRAMs

by

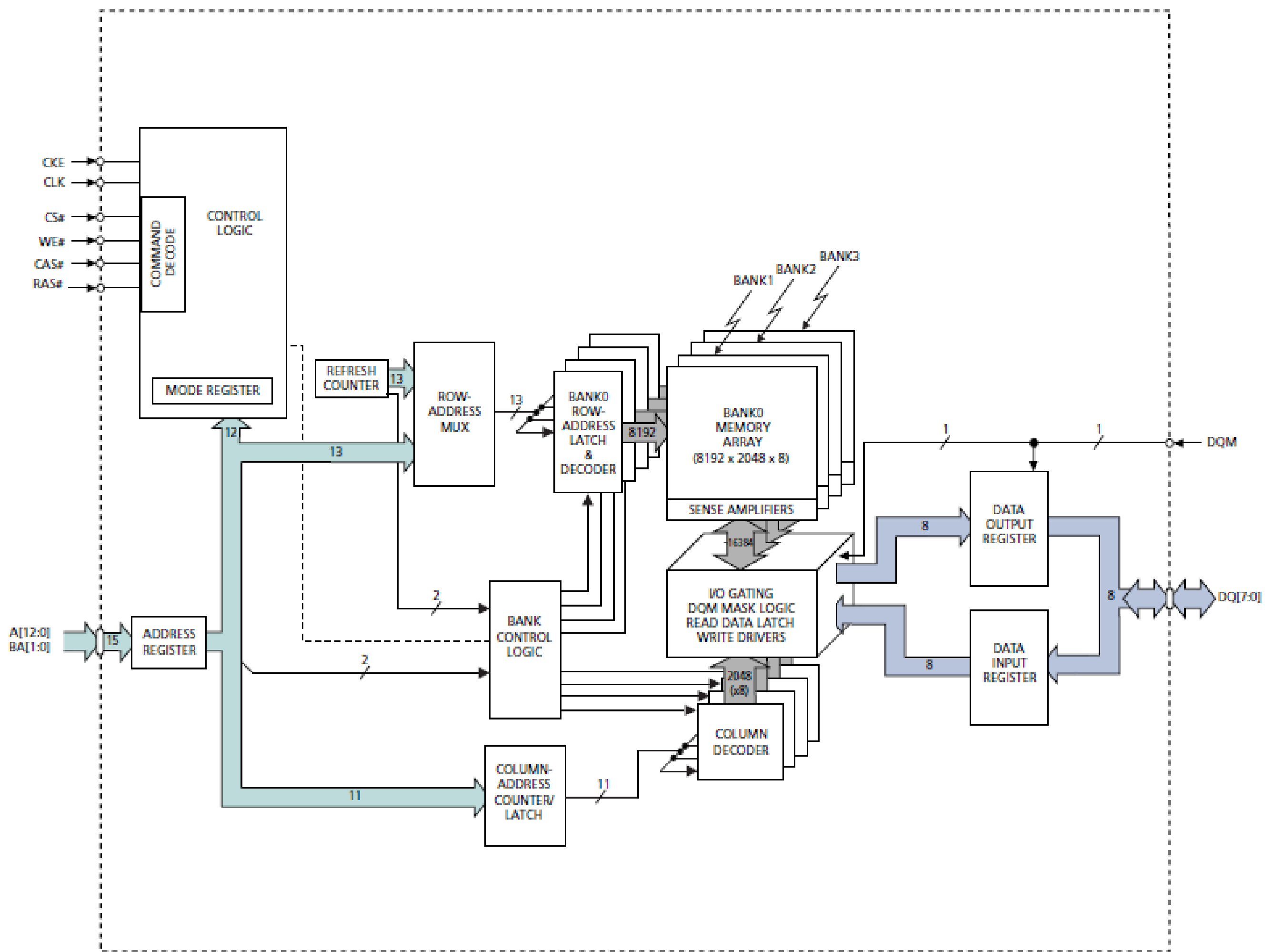
Dr. William R. Huber, P.E.

 Copyright© 2022 William R. Huber

A. Introduction

To understand the operation and advantages of Double Data Rate (DDR) Synchronous Dynamic Random Access Memories (DDR SDRAMs), we will start by reviewing the architecture of a single data rate (SDR) SDRAM. In keeping with industry practice, the balance of this course will refer to SDR SDRAMs simply as SDRAMs. A Micron data sheet provides a functional block diagram.[1]

Figure 2: 64 Meg x 8 Functional Block Diagram



[1] Micron MT48LC128M4A2 512Mb SDRAM Data Sheet, Rev. M 6/10 EN, page 8

 Copyright© 2022 William R. Huber

SDRAM architecture is very similar to that of an asynchronous DRAM, with the addition of Command Inputs (CS#, WE#, CAS# and RAS#)[2], Clock Inputs (CKE, CLK) and Control Logic including Command Decode and Mode Register. Some of the significant differences in the operation of an SDRAM as compared to a DRAM are as follows:

- SDRAM operation is initiated by the low to high transition of the external clock, and data bits are input or output in synchronism with the external clock;
- In an SDRAM, the operational mode (ACTIVE, READ, WRITE, REFRESH, etc.) is established by a command issued at the beginning of the cycle;
- The SDRAM command is defined by the logic state of several inputs (e.g.; RAS#, CAS#, WE# and A10) when CS# (Chip Select) is low, CKE (Clock Enable) is high, and CLK (external Clock) transitions from low to high;
- The SDRAM is organized into banks, which can operate virtually independently of each other. Thus, one bank can be providing data during a READ operation while another bank is performing a precharge or REFRESH operation; and
- Specific characteristics of the SDRAM operation, such as how many bits of data will be read out sequentially (burst length) and how many clock cycles will elapse between the READ command and the appearance of the first bit of data (latency), are specified by the system and stored in the SDRAM mode register prior to the first ACTIVE command.

B. <u>Double Data Rate (DDR) SDRAMs</u>

1. Comparison to SDRAM

From a user standpoint, the most apparent difference between single data rate and double data rate SDRAMs is the rate at which data is written to or read from the memory. As the name implies, DDR devices accept data from or

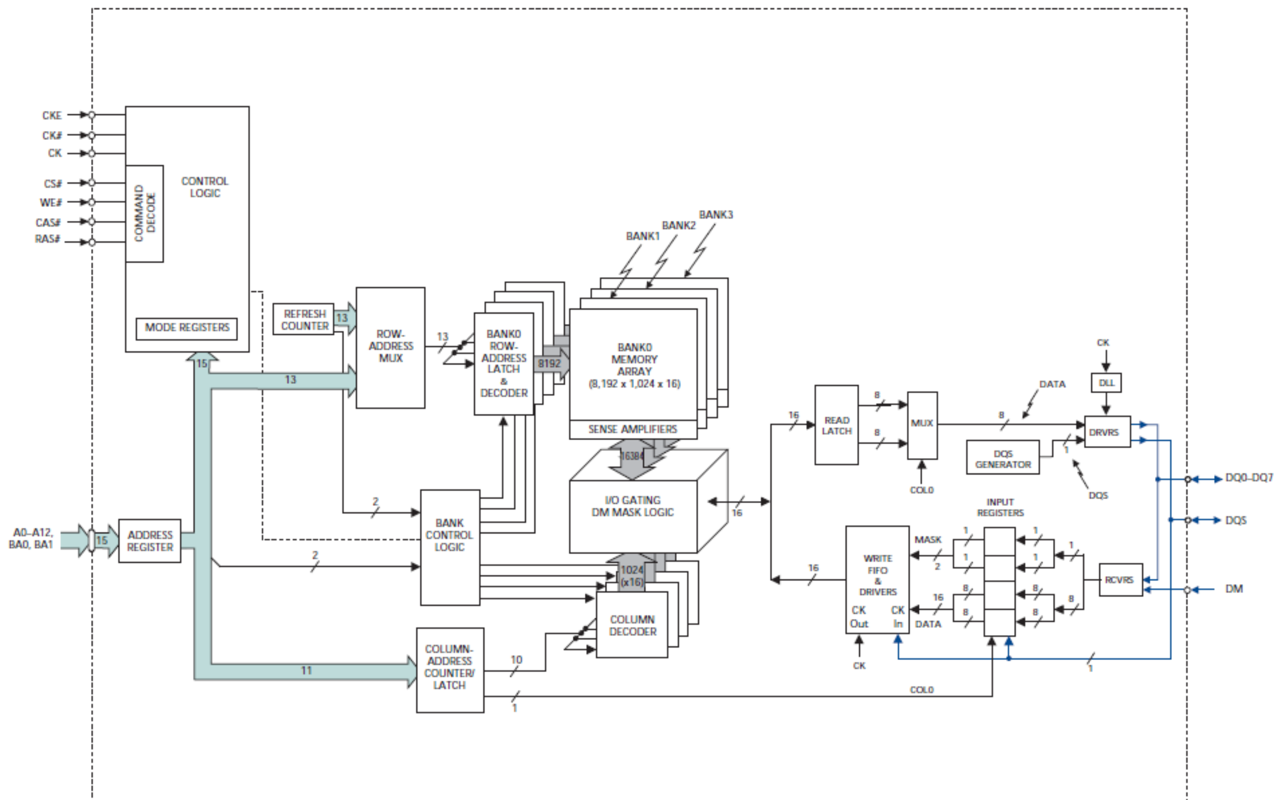
[2] The # following the letter symbol means that the signal is "active low." A /, either before or after the letter symbol, carries the same meaning.

 Copyright© 2022 William R. Huber

provide data to the bus twice as fast as SDR devices operating at the same external clock speed.

Now look at the functional block diagram for a Micron DDR SDRAM part:[3]

Figure 4: 64 Meg x 8 Functional Block Diagram



What are the differences between the SDR and DDR block diagrams?

- At the top left, the DDR device has two clock inputs, CK# and CK, where the SDR device has a single clock input, CLK.
- In the CONTROL LOGIC block at the upper left, the DDR device has MODE REGISTERS, where the SDR device has a MODE REGISTER.
- At the lower right, the DDR block diagram provides more detail for the output (READ) and input (WRITE) data paths than does the SDR diagram.

---

[3] Micron MT46V128M4 512Mb DDR SDRAM Data Sheet, Rev. B 2/09 EN, page 8

 Copyright© 2022 William R. Huber

However, the essential difference is in the tiny box called DLL in the DDR diagram. The SDR diagram has no such box or function. DLL stands for Delay Locked Loop, and we will examine this box in detail later. The delay locked loop is added to generate the precise internal clock signals required to meet the very stringent timing specifications of a DDR device. For example, in a DDR device the initial data in response to a READ command must appear within ±0.75ns (that is ±0.75 billionths of a second) of a clock transition. By way of comparison, in a comparable SDR device the same data must appear within 5.4ns after a clock transition.

We will now examine each of the differences in DDR devices to see why the change has been made and what features it provides.

2. CK and CK# Inputs

CK and CK# are nominally exact complements. For the Micron device discussed above (2.5V nominal power supply voltage), the voltage at which CK and CK# are required to cross is 1.25±0.2V.

Clock cycle time requirements depend on CAS Latency (discussed in a later section) and speed grade of the specific device. Speed grade refers to a set of timing requirements which are evaluated by the vendor at final package test. Fast speed grades are more desirable, and command a higher selling price. Clock cycle time specifications for the Micron device discussed above are summarized in the following table.

Clock Cycle Time, tCK

Speed Grade	-5B		-6, -6T		-75E, -75Z		-75	
CAS Latency	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
3	5 ns	7.5 ns	---	---	---	---	---	---
2.5	6 ns	13 ns	6 ns	13 ns	7.5 ns	13 ns	7.5 ns	13 ns
2	7.5 ns	13 ns	7.5 ns	13 ns	7.5 ns	13 ns	10 ns	13 ns

The corresponding maximum memory clock frequencies are:

5 ns→200 MHz (DDR-400) 7.5 ns→133 MHz (DDR-266)
6 ns→166 MHz (DDR-333) 10 ns→100 MHz (DDR-200)

 Copyright© 2022 William R. Huber

Thus, a DDR SDRAM with a minimum clock cycle time of 5 ns can work in a system with a memory clock frequency of 200 MHz. Because it delivers 2 bits of information per clock cycle per data pin, such a device provides 400 Mb per second per data pin, and is referred to as a DDR-400. If the device is organized to have 8 data pins (data-in/data out, or DQ), it can output data at 3200 Mb per second (400 MB/s).[4]

3. Command Decoder

In a DDR SDRAM, a set of 4 signals (CS#, RAS#, CAS# and WE#) is controlled to indicate what operation is to be performed. The table below[5] shows the possibilities:

Table 28: Truth Table 1 – Commands

CKE is HIGH for all commands shown except SELF REFRESH; All states and sequences not shown are illegal or reserved

Function	CS#	RAS#	CAS#	WE#	Address	Notes
DESELECT	H	X	X	X	X	1
NO OPERATION (NOP)	L	H	H	H	X	1
ACTIVE (select bank and activate row)	L	L	H	H	Bank/row	2
READ (select bank and column and start READ burst)	L	H	L	H	Bank/col	3
WRITE (select bank and column and start WRITE burst)	L	H	L	L	Bank/col	3
BURST TERMINATE	L	H	H	L	X	4
PRECHARGE (deactivate row in bank or banks)	L	L	H	L	Code	5
AUTO REFRESH or SELF REFRESH (enter self refresh mode)	L	L	L	H	X	6, 7
LOAD MODE REGISTER	L	L	L	L	Op-code	8

Notes:
1. DESELECT and NOP are functionally interchangeable.
2. BA0–BA1 provide bank address and A0–A*n* (128Mb: *n* = 11; 256Mb and 512Mb: *n* = 12; 1Gb: *n* = 13) provide row address.
3. BA0–BA1 provide bank address; A0–A*i* provide column address, (where A*i* is the most significant column address bit for a given density and configuration, see Table 2 on page 2) A10 HIGH enables the auto precharge feature (non persistent), and A10 LOW disables the auto precharge feature.
4. Applies only to READ bursts with auto precharge disabled; this command is undefined (and should not be used) for READ bursts with auto precharge enabled and for WRITE bursts.
5. A10 LOW: BA0–BA1 determine which bank is precharged. A10 HIGH: all banks are precharged and BA0–BA1 are "Don't Care."
6. This command is AUTO REFRESH if CKE is HIGH; SELF REFRESH if CKE is LOW.
7. Internal refresh counter controls row addressing while in self refresh mode, all inputs and I/Os are "Don't Care" except for CKE.
8. BA0–BA1 select either the mode register or the extended mode register (BA0 = 0, BA1 = 0 select the mode register; BA0 = 1, BA1 = 0 select extended mode register; other combinations of BA0–BA1 are reserved). A0–A*n* provide the op-code to be written to the selected mode register.

---

[4] Mb stands for megabits (million bits); MB stands for megabytes (million bytes, where 1 byte = 8 bits)
[5] Micron, op. cit., page 43

 Copyright© 2022 William R. Huber

We will examine three of the possible operations: ACTIVE, READ and LOAD MODE REGISTER.

a. ACTIVE Operation

To obtain information stored in any DRAM, it is necessary to select the row and column where the desired information is stored. The ACTIVE operation accomplishes the first part of this selection by identifying the desired row. To command an ACTIVE operation, CS# and RAS# are set to a logic low; CAS# and WE# are set high; the desired bank (of the 4 banks available) is identified by bank address inputs BA0 and BA1; and the desired row (of the 8192 available) is identified by binary address inputs A0 to A12 (2^{13} = 8192). After all of these inputs are set, they are registered (detected and stored) on-chip by the next rising edge of clock CK.

Here are some of the actions that occur inside the DDR SDRAM during the ACTIVE operation:

- Command inputs, bank addresses and row addresses are detected, amplified and stored in corresponding registers.
- Command inputs are decoded to determine what operation is to be performed.
- Bank addresses select the desired bank and steer the row addresses to the row decoders in that bank.
- Row addresses are decoded from their binary representations to select the single desired row in the selected bank.
- The selected row (word) line is turned on (usually driven to a positive voltage), thus connecting every memory cell along that row to its corresponding bit line.
- The stored charge (or lack of charge) in each of the accessed memory cells slightly increases or decreases the voltages on the associated bit lines.
- Sense amplifiers connected to each of the bit lines detect the change in voltages on the bit lines by comparing the bit line voltages to a reference voltage, and then amplify that voltage difference.
- The sense amplifiers send the amplified voltages back along the bit lines to the selected memory cells, thus refreshing the cells to their original states.
- The sense amplifiers store the information from the selected memory cells for further operations.

 Copyright© 2022 William R. Huber

b. READ Operation

The READ operation accomplishes the second part of the row and column selection by identifying the desired starting column.[6] To command a READ operation, CS# and CAS# are set to a logic low; RAS# and WE# are set high; the desired bank (of the 4 banks available) is identified by bank address inputs BA0 and BA1; the desired starting column (of the 2048 available) is identified by binary address inputs A0 to A9 and A11 ($2^{11} = 2048$); and address input A10 is set either high or low, as desired.[7] After all of these inputs are set, they are registered (detected and stored) on-chip by the next rising edge of clock CK.

Here are some of the actions that occur inside the DDR SDRAM during the READ operation:

- Command inputs, bank addresses and column addresses are detected, amplified and stored in corresponding registers.
- Command inputs are decoded to determine what operation is to be performed.
- Bank addresses select the desired bank and steer the column addresses to the column decoders in that bank.
- Column addresses are decoded from their binary representations to select the starting column and the next 15 columns[8] in the selected bank.
- Column select signals are sent to the transmission gates (switches) in each of the selected columns, thus connecting the outputs of the sense amplifiers in those selected columns to internal input/output (I/O) lines.
- Data on the I/O lines is amplified, and data from the selected column and the next 7 columns are sent to the output drivers.
- At the CK-CK# crossing defining the end of the CAS Latency time, the output drivers are activated and provide data on data pins DQ0 to DQ7.

---

[6] The column address input identifies the desired **starting** column for reading and outputting information. Because this memory device has 8 input/output (DQ) pins (referred to as x8 pinout), data from the addressed column will appear on DQ0; and data from the next 7 columns in sequence will appear simultaneously on DQ1 to DQ7 respectively.

[7] A10 high enables Auto-Precharge, thus returning the selected bank to the idle state after the READ operation is complete. A10 low disables Auto-Precharge, thus keeping the bank, row and column selected pending future operations.

[8] Depending on the specific chip architecture, 8, 16 or 32 columns will be selected at this time. For a x4 pinout, 8 columns are selected; for x8, 16 columns are selected; and for x16, 32 columns are selected. This is referred to as 2n-bit prefetch; two outputs worth of bits are selected and processed so that the information for two consecutive outputs, on the rising and falling clock edges, will be available.

 Copyright© 2022 William R. Huber

- Depending on the Burst Length stored in the Mode Register,[9] more data in 8-bit bursts appears on DQ0 to DQ7 at the next several crossings of CK and CK#.

c. Load Mode Register (LMR) Operation

The Mode Registers are specialized on-chip static memory structures to store operating parameters for the DDR SDRAM. A DDR SDRAM has two Mode Registers. The first, simply called Mode Register, is very similar to that of an SDRAM. The second, called the Extended Mode Register, provides storage to define additional functions. First we will examine the Mode Register.[10]

Mode Register

The mode register is used to define the specific DDR SDRAM mode of operation. This definition includes the selection of a burst length, a burst type, a CAS latency, and an operating mode, as shown in Figure 23. The mode register is programmed via the LMR command (with BA0 = 0 and BA1 = 0) and will retain the stored information until it is programmed again or until the device loses power (except for bit A8, which is self-clearing).

Reprogramming the mode register will not alter the contents of the memory, provided it is performed correctly. The mode register must be loaded (reloaded) when all banks are idle and no bursts are in progress, and the controller must wait the specified time before initiating the subsequent operation. Violating either of these requirements will result in unspecified operation.

Mode register bits A0–A2 specify the burst length, A3 specifies the type of burst (sequential or interleaved), A4–A6 specify the CAS latency, and A7–A*n* specify the operating mode.

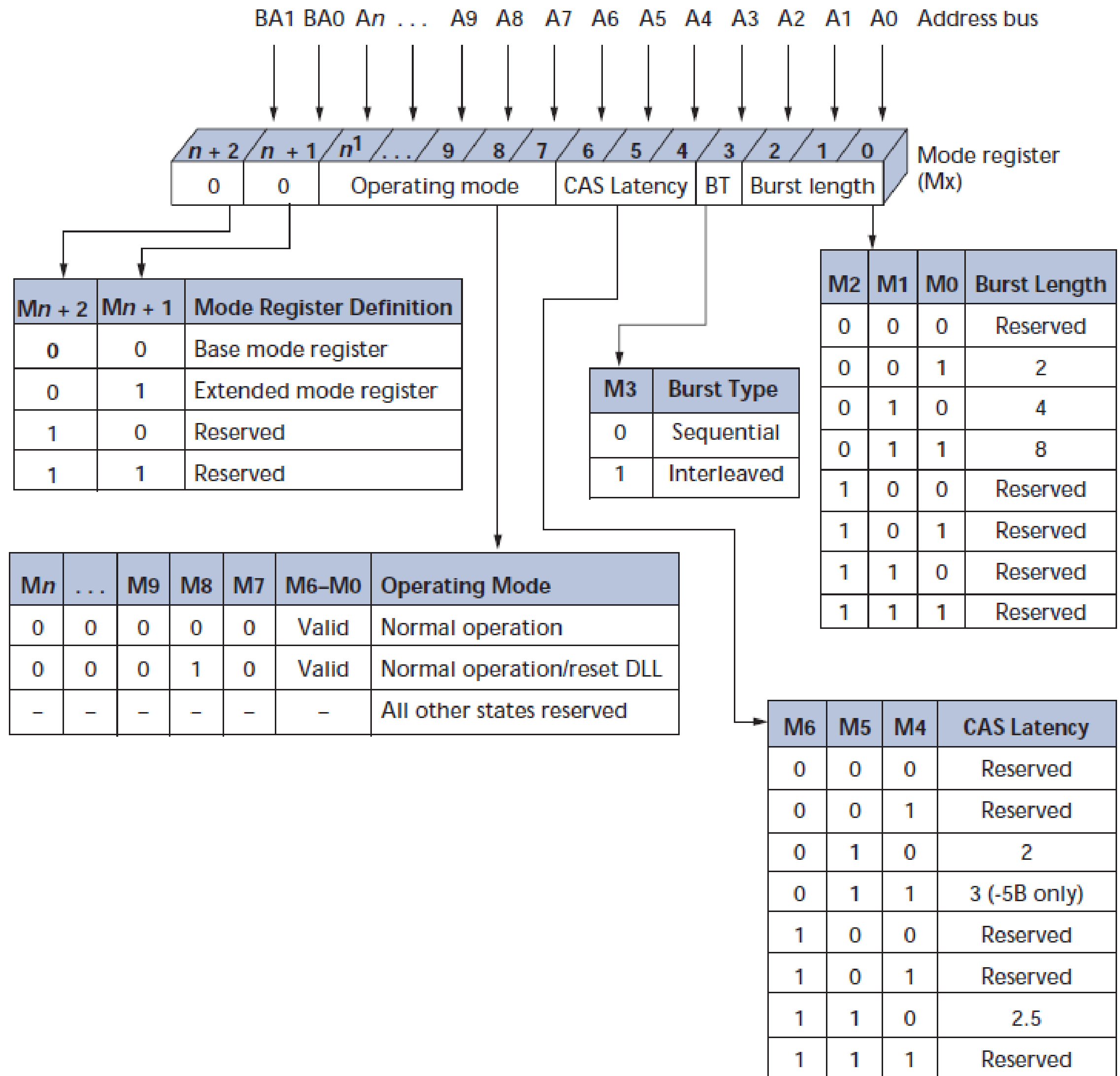
To command a Load Mode Register (LMR) operation, CS#, RAS#, CAS# and WE# are all set low; bank addresses BA0 and BA1 are set low; and address inputs A0 to A12 are set to carry the information to be loaded into the 13 bits of the Mode Register (A0 to the first bit, etc.). On the next rising edge of clock CK, the information is loaded into and stored by the Mode Register. The LMR operation is typically done when the system is booting up, but can be done any time all memory banks are idle and no data bursts are in progress.

---

[9] Burst Length and Mode Register are discussed in the next section.
[10] Micron MT46V128M4 512Mb DDR SDRAM Data Sheet, Rev. B 2/09 EN, page 55

 Copyright© 2022 William R. Huber

Figure 23: Mode Register Definition



Mn + 2	Mn + 1	Mode Register Definition
0	0	Base mode register
0	1	Extended mode register
1	0	Reserved
1	1	Reserved

M3	Burst Type
0	Sequential
1	Interleaved

M2	M1	M0	Burst Length
0	0	0	Reserved
0	0	1	2
0	1	0	4
0	1	1	8
1	0	0	Reserved
1	0	1	Reserved
1	1	0	Reserved
1	1	1	Reserved

Mn	. . .	M9	M8	M7	M6–M0	Operating Mode
0	0	0	0	0	Valid	Normal operation
0	0	0	1	0	Valid	Normal operation/reset DLL
–	–	–	–	–	–	All other states reserved

M6	M5	M4	CAS Latency
0	0	0	Reserved
0	0	1	Reserved
0	1	0	2
0	1	1	3 (-5B only)
1	0	0	Reserved
1	0	1	Reserved
1	1	0	2.5
1	1	1	Reserved

Notes: 1. *n* is the most significant row address bit from Table 2 on page 2. For this device, *n* = 12.

Copyright© 2022 William R. Huber

(1) Burst Length

Mode Register bits 0, 1 and 2 store burst length information. Burst length defines how many sequential bits of data will appear on each DQ pin during a READ operation. For example, if Mode Register bits 2-1-0 store 0-1-0, then a burst length of 4 is called for. For each READ operation, assuming a x8 device type, 8 bits of data will appear on pins DQ0 to DQ7 respectively at the end of the CAS Latency time. Then one-half clock cycle later, at the next crossing of clocks CK and CK#, another set of 8 bits will appear. For a burst length of 4, 4 sets of 8 bits each will appear on pins DQ0 to DQ7. Both READ and WRITE operations are burst oriented.

(2) Burst Type

Mode Register bit 3 stores burst type information. Burst type can be either Sequential or Interleaved. The ordering of accesses within a burst is determined by the burst length, the burst type and the starting column address, as summarized in the following table.[11]

Table 34: Burst Definition

Burst Length	Starting Column Address			Order of Accesses Within a Burst	
				Type = Sequential	Type = Interleaved
2	–	–	**A0**	–	–
	–	–	0	0-1	0-1
	–	–	1	1-0	1-0
4	–	**A1**	**A0**	–	–
	–	0	0	0-1-2-3	0-1-2-3
	–	0	1	1-2-3-0	1-0-3-2
	–	1	0	2-3-0-1	2-3-0-1
	–	1	1	3-0-1-2	3-2-1-0
8	**A2**	**A1**	**A0**	–	–
	0	0	0	0-1-2-3-4-5-6-7	0-1-2-3-4-5-6-7
	0	0	1	1-2-3-4-5-6-7-0	1-0-3-2-5-4-7-6
	0	1	0	2-3-4-5-6-7-0-1	2-3-0-1-6-7-4-5
	0	1	1	3-4-5-6-7-0-1-2	3-2-1-0-7-6-5-4
	1	0	0	4-5-6-7-0-1-2-3	4-5-6-7-0-1-2-3
	1	0	1	5-6-7-0-1-2-3-4	5-4-7-6-1-0-3-2
	1	1	0	6-7-0-1-2-3-4-5	6-7-4-5-2-3-0-1
	1	1	1	7-0-1-2-3-4-5-6	7-6-5-4-3-2-1-0

[11] Ibid., page 56

 Copyright© 2022 William R. Huber

(3) CAS Latency

Mode Register bits 4, 5 and 6 store CAS Latency (CL) information. CAS Latency defines the time from the registration of the READ command to the appearance of valid data on the Data Input/Output (DQ) pins, and is measured in clock cycles. Typical values of CAS Latency are 2, 2.5 and 3. From the Micron 512Mb DDR SDRAM Data Sheet:[12]

CAS Latency (CL)

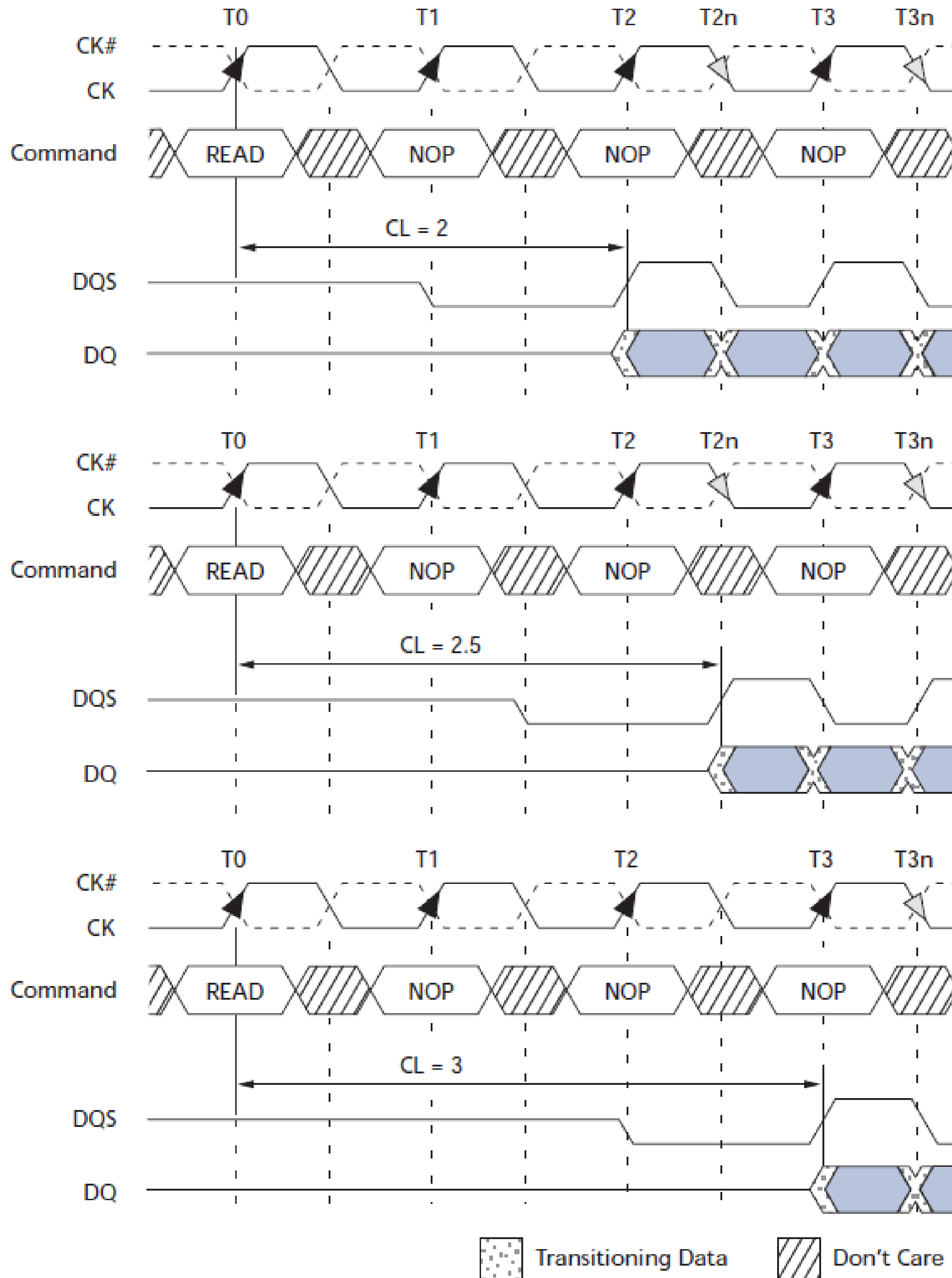
The CL is the delay, in clock cycles, between the registration of a READ command and the availability of the first bit of output data. The latency can be set to 2, 2.5, or 3 (-5B only) clocks, as shown in Figure 24. Reserved states should not be used, as unknown operation or incompatibility with future versions may result.

If a READ command is registered at clock edge *n*, and the latency is *m* clocks, the data will be available nominally coincident with clock edge $n + m$. Table 35 on page 58 indicates the operating frequencies at which each CL setting can be used.

[12] Ibid., p. 57

 Copyright© 2022 William R. Huber

The following waveforms illustrate CL values of 2, 2.5 and 3 for a READ operation. Note that data (DQ) appears almost exactly at the end of the CAS Latency time.



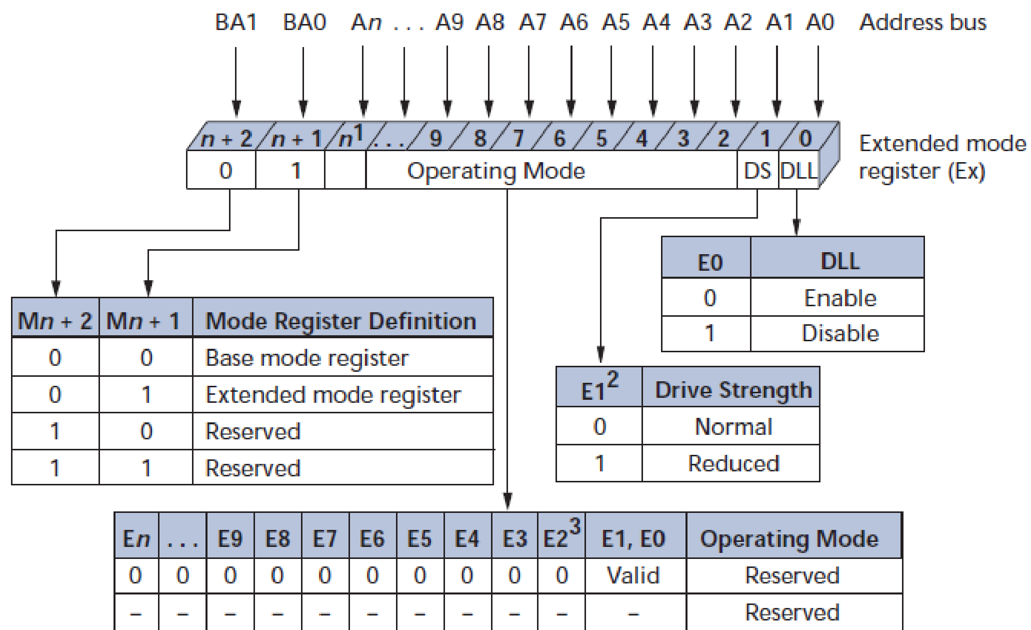
Copyright© 2022 William R. Huber

(4) Operating Mode

Mode Register bits 7 through 12 store Operating Mode information. In this DDR SDRAM there are only two valid operating modes, Normal operation and Normal operation/Reset DLL. If the memory clock frequency is changed during operation, most memories of this type require that the Delay Locked Loop (DLL) be reset.

Next we will examine the Extended Mode Register.[13] The Extended Mode Register provides storage for information determining additional operating options in a DDR SDRAM that are not available in an SDR SDRAM.

Figure 25: Extended Mode Register Definition



Notes: 1. *n* is the most significant row address bit from Table 2 on page 2.

(5) Drive Strength (DS) Options

Consider first the Drive Strength option which is specified by bit 1 of the Extended Mode Register. Either Normal or Reduced drive strength can be selected. For the Micron device we have been discussing, Normal drive strength corresponds to an output current drive of at least ±16.8 mA (for output low and output high). Reduced drive strength corresponds to an output current drive of at least ±9 mA.

---

[13] Ibid., p. 59

Copyright© 2022 William R. Huber

(6) DLL Options

Now consider the DLL option specified by bit 0 of the Extended Mode Register. The DLL can be either Enabled or Disabled. For normal operation, and for virtually all of the device data sheet specifications to apply, the DLL must be enabled. The DLL can be disabled during Power-down operation (to reduce power consumption) and when running with a lower-than-specified memory clock frequency (possibly during system debugging).

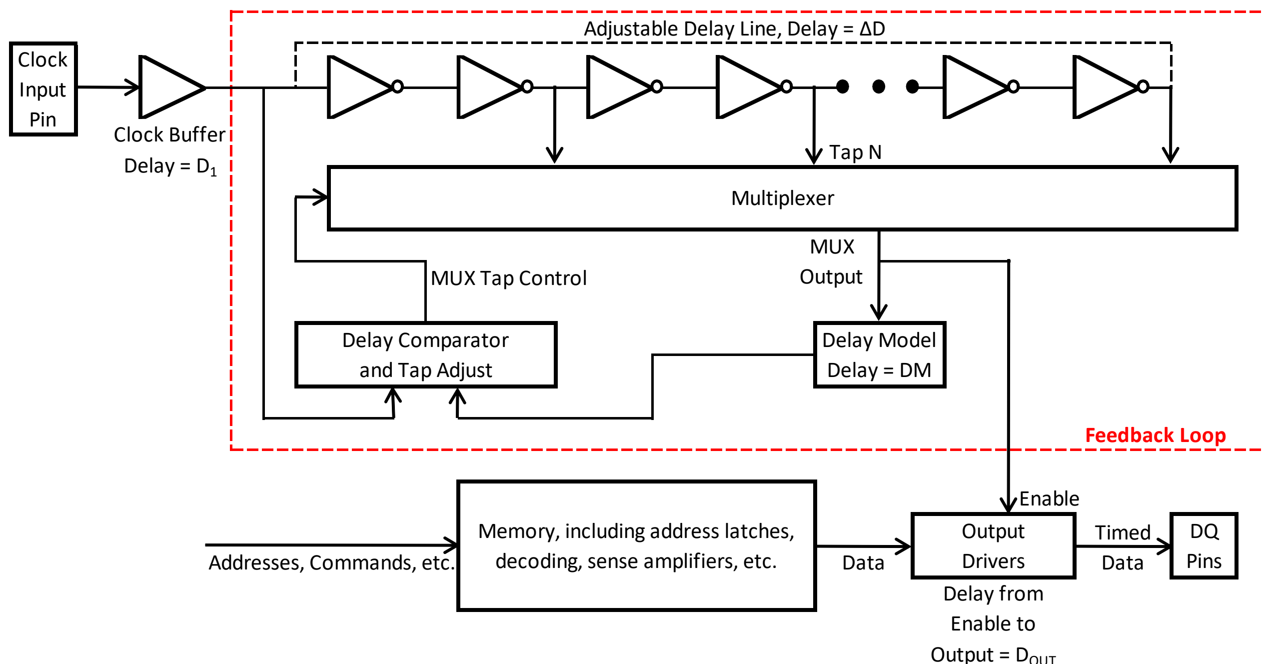
4. Delay Locked Loop (DLL)

a. Basic Function and Block Diagram

The primary purpose of the DLL in DDR SDRAMs is to align the data edges to the clock transitions at the end of the CAS Latency. As shown in the waveforms on page 13, the first data bit appears coincident with the clock crossing defining the end of CAS Latency; and successive data bits start exactly on successive clock crossings. While nothing is exact, the timing tolerance between the clock crossings and the data valid times is on the order of ± 0.75 ns, a very tight window. Because circuit delays on the chip vary with device processing, temperature and power supply voltage, a circuit such as a DLL is required to meet the rigid timing requirements. While the on-chip DLL is very complex, we can understand the operation with a simple example.[14]

[14] Adapted from Foss, R.C. et al; "Delay Locked Loop Implementation in a Synchronous Dynamic Access Memory;" U. S. Patent 6,205,083; Issued March 20, 2001; Figure 5

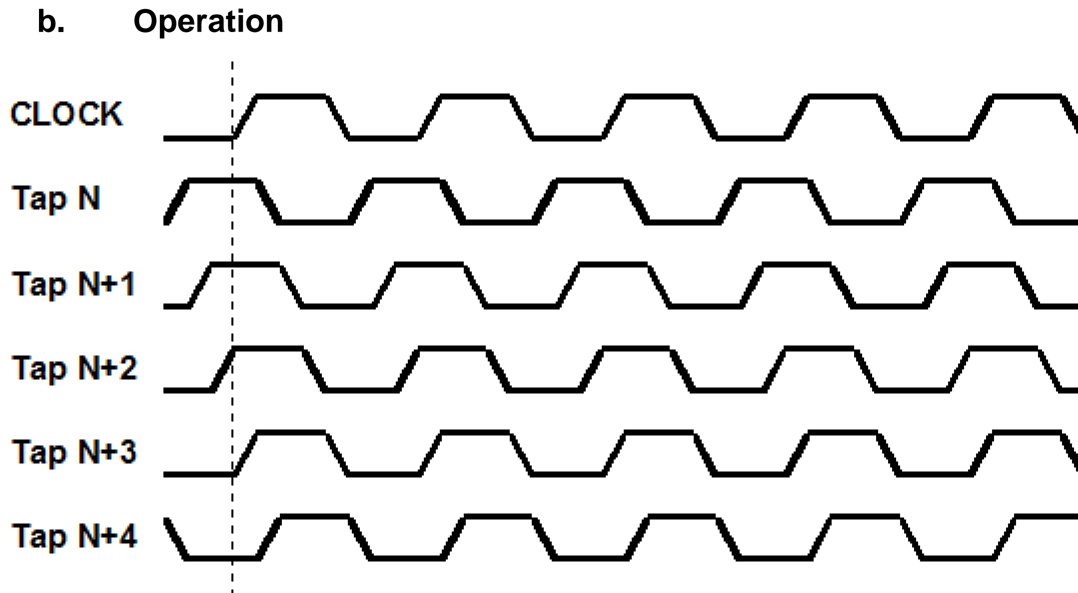
 Copyright© 2022 William R. Huber



The external clock enters at the top left, on the Clock Input Pin. To isolate the external clock from high capacitive loads on-chip, it immediately enters a Clock Buffer. Unfortunately, the Clock Buffer adds delay, D_1, to the on-chip clock signal. The internal clock then goes to two circuits: an Adjustable Delay Line composed of cascaded inverters; and a Delay Comparator and Tap Adjust circuit. The Adjustable Delay Line has an output tap after each pair of inverters. These taps go to an N-to-1 Multiplexer, the output of which is controlled by the output of the Delay Comparator and Tap Adjust circuit. The Multiplexer output also goes to two circuits: the Delay Model with delay = D_M and the Enable terminal of the Output Drivers. The Delay Model is inserted to compensate for the delays of the Clock Buffer and the Output Drivers, as we shall see momentarily. The Output Drivers provide an amplifying and buffering function for the data from the memory so it can be driven onto the external data bus. Finally, the Delay Comparator and Tap Adjust circuit compares its two inputs and adjusts the Multiplexer tap until those two inputs rise simultaneously.

Copyright© 2022 William R. Huber

b. Operation



Signals in Adjustable Delay Line

Consider the waveforms shown in the figure above. **CLOCK** is the internal clock signal (after the Clock Buffer) and is the input to the Adjustable Delay Line and to one input of the Delay Comparator and Tap Adjust circuit. The waveform labeled **Tap N** is from one of the middle inverters of the delay line. The waveform labeled **Tap N+1** is from the next delay line tap, etc. Suppose that the initial tap selected by the MUX is Tap N. Thus, the waveform from **Tap N** is fed to the Delay Comparator and Tap Adjust circuit, and this circuit determines that the rising edge of the **Tap N** signal occurs before the rising edge of the **CLOCK** signal. Therefore, the circuit instructs the MUX to select a later tap, **Tap N+1**. However, when the Delay Comparator and Tap Adjust circuit compares the **Tap N+1** waveform to the **CLOCK** waveform, it determines that the rising edge of the **Tap N+1** waveform still occurs first. After selecting and comparing the waveform from **Tap N+2** to the **CLOCK**, the Delay Comparator and Tap Adjust circuit finally instructs the MUX to select **Tap N+3**. Now the Delay Comparator and Tap Adjust circuit finds that the rising edges of the waveform from **Tap N+3** and the **CLOCK** are simultaneous, so no further tap adjustment is needed. This condition is termed "DLL Locked", meaning that the correct tap has been found. If the MUX had selected the waveform from **Tap N+4**, the Delay Comparator and Tap Adjust circuit would have found that the rising edge of that waveform occurs <u>after</u> the rising edge of **CLOCK** and would have instructed the MUX to select **Tap N+3**.

Copyright© 2022 William R. Huber

c. Timing Analysis

The objective of the DLL is to cause the data to become valid at the rising edge of the Clock Input signal. Therefore, depending on whether the CAS Latency is an integer or an integer + ½, the total delay from the Clock Input Pin to the DQ Pins must be a whole number of clock cycles or a whole number + ½ cycle. Thus,

$$D_1 + \Delta D + D_{OUT} = N/2 \times T_{CY} \quad (1)$$

where N is a positive integer and T_{CY} is the clock cycle time.

When the DLL is in the locked mode (i.e., under normal operation), the two inputs to the Delay Comparator and Tap Adjust block will both rise (and fall) simultaneously. Therefore:

$$D_1 + N/2 \times T_{CY} = D_1 + \Delta D + D_M \quad (2)$$

Solving equations (1) and (2) for T_{CY} and equating the results:

$$2/N \times (D_1 + \Delta D + D_{OUT}) = 2/N \times (\Delta D + D_M) \quad (3)$$

Solving equation (3) for D_M:

$$\boxed{D_M = D_1 + D_{OUT}} \quad (4)$$

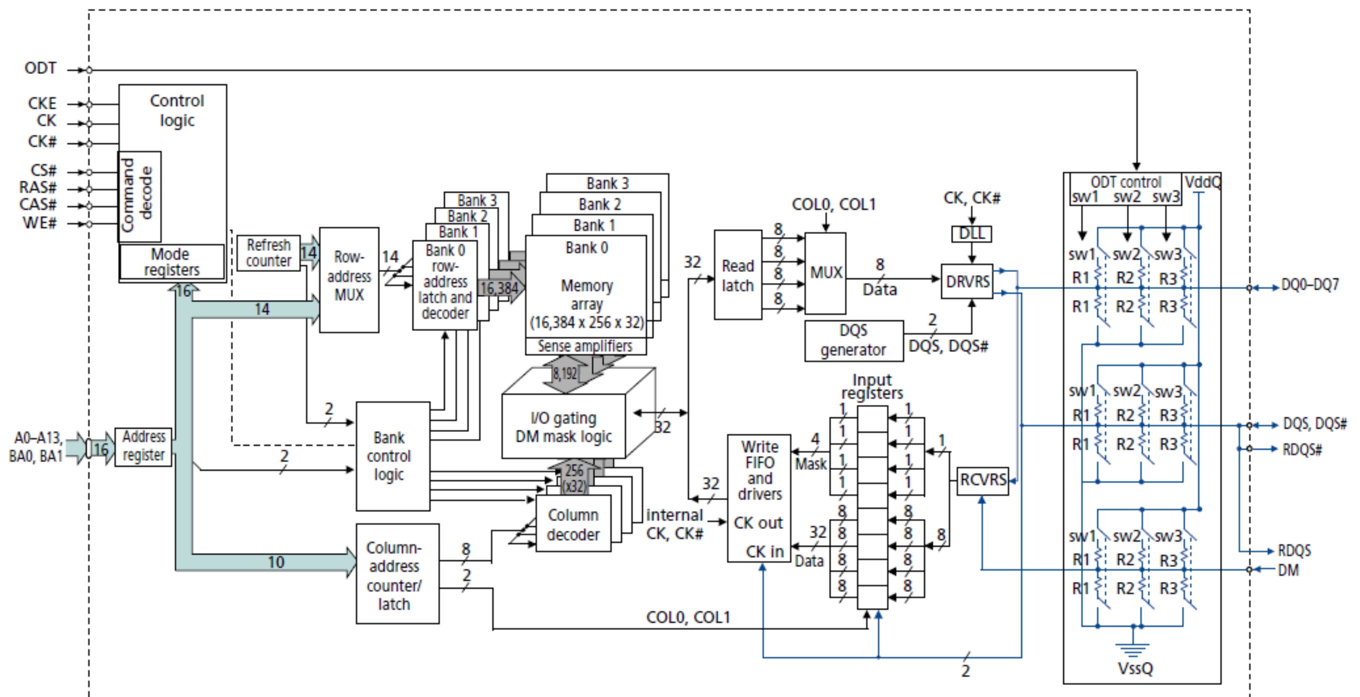
So, to satisfy the requirement that data becomes valid at the clock edge, the delay of the Delay Model must be equal to the sum of the Clock Buffer Delay and the Output Driver Delay from Enable to Output. More generally, the Delay Model delay must equal the sum of all delays from the Clock Input Pin to the Data Output Pins that lie **outside** of the Feedback Loop.

 Copyright© 2022 William R. Huber

C. Double Data Rate II (DDR2) SDRAMs

The DDR2 architecture was introduced in mid-2003 and, according to Micron,[15] was the dominant DRAM architecture until the end of 2009. Let's start by looking at a functional block diagram of a Micron 512Mb DDR2 SDRAM:[16]

Figure 4: 64 Meg x 8 Functional Block Diagram



Comparing this block diagram to that for the DDR SDRAM on page 4 reveals one major difference: the addition of the ODT (On-Die Termination) block at the far right. We will explore the ODT function next, and then discuss some of the other differences between DDR and DDR2 SDRAMs.

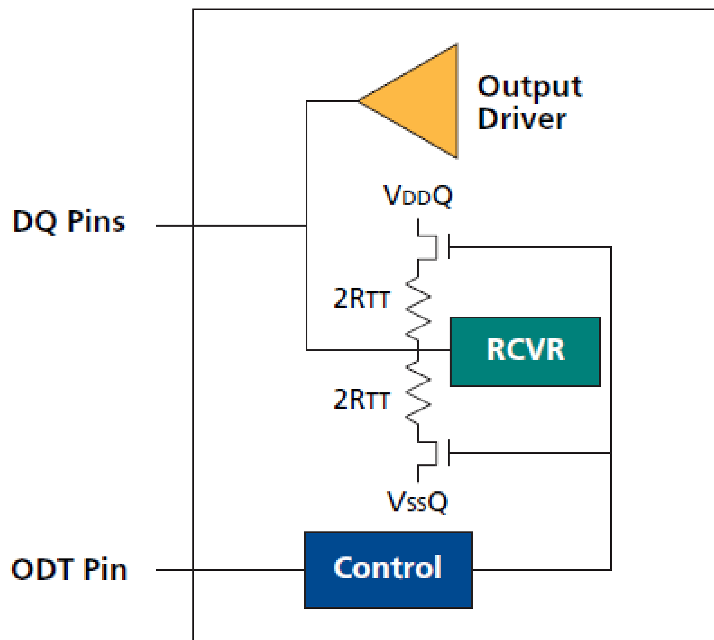
1. ODT Capability

At the high data bus frequencies (200 MHz and higher) used in modern systems, the bus conductors act as transmission lines and must be terminated in their characteristic impedance to minimize ringing and the resulting signal degradation. DDR2 SDRAMs provide that termination on-chip, thus reducing the external components needed. In addition, the on-chip termination is programmable, both as to

[15] "DDR3 Advantages," Micron Marketing, 4/8/2009
[16] Micron MT47H64M8 512Mb DDR2 SDRAM Data Sheet, Rev. Q 10/10 EN, page 12

Copyright© 2022 William R. Huber

the value of the resistors and whether or not they are connected for any given operation. The following diagram shows how the on-die termination looks.[17]



Each Data In/Data Out (DQ) pin has two resistors (shown as 2RTT) attached to it. These resistors can be connected, under program control, to the power supply voltage (VDDQ) and ground (VSSQ). Under normal conditions, the termination is activated (transistors turned on) during WRITE operations to the memory chip; and deactivated (transistors turned off) during READ operations from the memory chip.

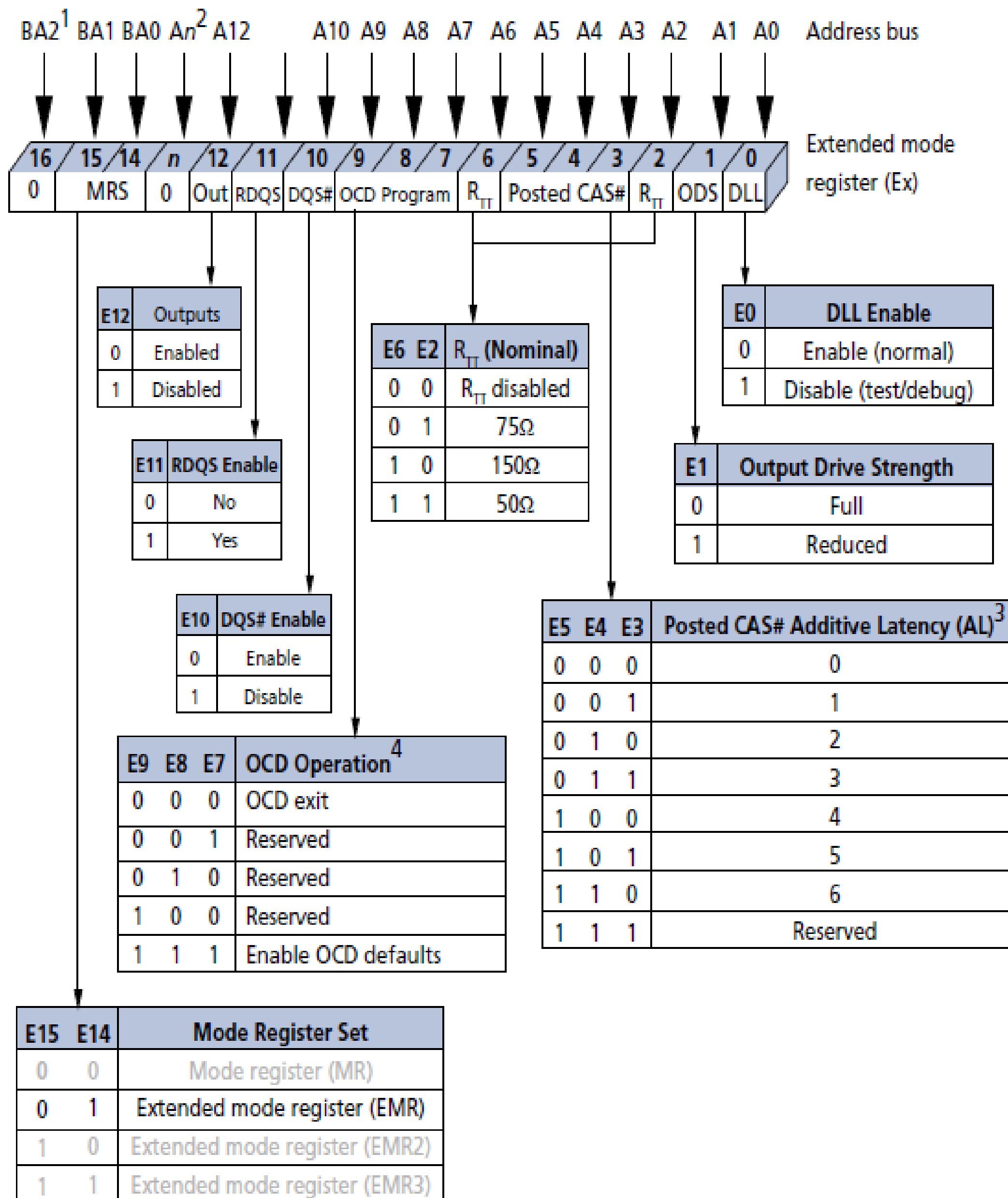
The values of the resistors are determined by bits in the Extended Mode Register. Whether or not the transistors are turned on, thus connecting the resistors to their corresponding power supplies, is controlled by the state of the ODT pin when the command is registered (on the rising edge of CK).

2. Extended Mode Register

The functions controlled by the Extended Mode Register are greatly expanded in the DDR2 architecture. The following diagram[18] provides some details:

[17] Micron TN-47-02 “DDR2 Offers New Features/Functionality,” page 3
[18] Micron MT47H64M8 512Mb DDR2 SDRAM Data Sheet, Rev. Q 10/10 EN, page 79

 Copyright© 2022 William R. Huber



E12	Outputs
0	Enabled
1	Disabled

E11	RDQS Enable
0	No
1	Yes

E10	DQS# Enable
0	Enable
1	Disable

E9	E8	E7	OCD Operation[4]
0	0	0	OCD exit
0	0	1	Reserved
0	1	0	Reserved
1	0	0	Reserved
1	1	1	Enable OCD defaults

E6	E2	R_{TT} (Nominal)
0	0	R_{TT} disabled
0	1	75Ω
1	0	150Ω
1	1	50Ω

E0	DLL Enable
0	Enable (normal)
1	Disable (test/debug)

E1	Output Drive Strength
0	Full
1	Reduced

E5	E4	E3	Posted CAS# Additive Latency (AL)[3]
0	0	0	0
0	0	1	1
0	1	0	2
0	1	1	3
1	0	0	4
1	0	1	5
1	1	0	6
1	1	1	Reserved

E15	E14	Mode Register Set
0	0	Mode register (MR)
0	1	Extended mode register (EMR)
1	0	Extended mode register (EMR2)
1	1	Extended mode register (EMR3)

As with the DDR SDRAM, this DDR2 Extended Mode Register includes bits to control the DDR Enable (bit 0) and the Output Drive Strength (bit 1). But in the DDR2 architecture, bits 2 and 6 control the ODT function. 0-0 in those two bits disables the termination (i.e., turns off the transistors so the ends of the resistors are floating). Other bit values set the resistance values such that their parallel equivalent resistance is either 75, 150 or 50 ohms.

 Copyright© 2022 William R. Huber

3. Posted CAS# Capability

A new capability called Posted CAS# is added to the DDR2 architecture. To understand the advantage of Posted CAS#, we need to review the older DDR SDRAM architecture and the way data was read or written in that architecture. Reading (or writing) data in a DDR SDRAM requires two distinct operations: ACTIVE (to provide the bank and row address); and READ (or WRITE) (to provide the bank and starting column address. ACTIVE and READ (WRITE will be understood as another possible command in the following discussion) must be separated in time by the timing parameter t_{RCD}, which is specified in nanoseconds on each device data sheet. For the Micron DDR SDRAM we examined in Section B above, t_{RCD}(Min) is either 15 or 20 ns, depending on the speed grade.

Requiring two commands separated in time by a number of nanoseconds creates several problems for the memory controller, which issues commands to the memories in a system.

- First, the controller (and the software that runs it) must remember to issue the second command at the appropriate time after issuing the first command.
- Second, memory controllers measure time in clock cycles, not in numbers of nanoseconds.
- Third, the second command might not be able to be issued at the optimum time because another command is in that time slot (likely in multi-bank operation).

So it would be advantageous if the memory controller could issue the two commands on consecutive clock cycles and let the memory accept the second command (READ or WRITE) at a predetermined time after the first command (ACTIVE). Posted CAS# provides that capability.

With Posted CAS#, the READ command is issued in the very next time slot after the ACTIVE command. This READ command is accepted by the DDR2 SDRAM, but not executed until completion of the number of clock cycles specified as Additive Latency (AL) in the Posted CAS# bits of the Extended Mode Register. For the device we are discussing, AL can be any value from 0 to 6 clock cycles. However, the internal memory still has an inherent minimum time from the ACTIVE command, t_{RCD}(Min), before it can execute a READ command. So, assuming the READ command is issued in the next time slot after the ACTIVE command, it is a requirement that

$AL \geq t_{RCD}(Min) - 1$; where t_{RCD}(Min) is expressed in clock cycles.

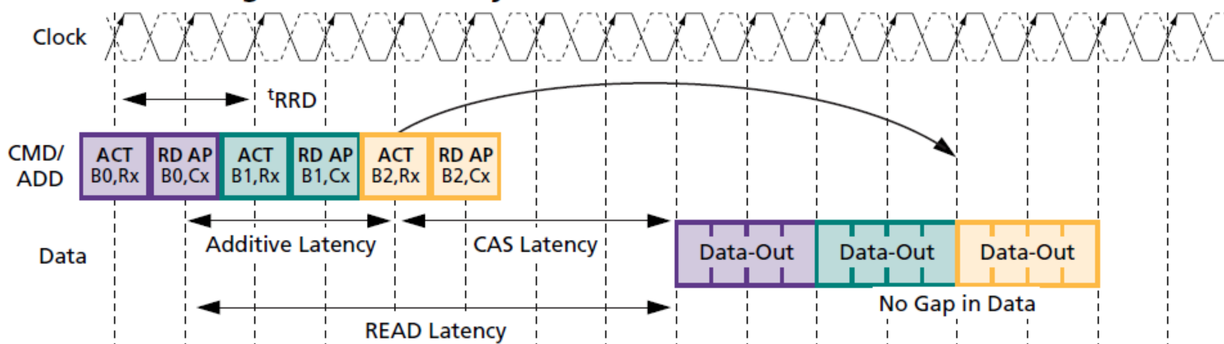
 Copyright© 2022 William R. Huber

For example: for the Micron MT47H64M8 512Mb DDR2 SDRAM we have been discussing, for the speed grade -25E, t_{RCD}(Min) is 12.5 ns and the minimum clock cycle time is 3.75 ns for CAS Latency (CL) = 4.[19] So

AL ≥ t_{RCD}(Min) – 1, or AL ≥ (12.5/3.75) – 1, or AL ≥ 3.33 – 1 = 2.33.

Therefore, AL would have to be specified as 3 or greater for this set of operating conditions.

The following timing diagram shows how Posted CAS# and multi-bank accessing allow a continuous data stream, with no gaps. For this example, AL = 3, CL = 4 and Burst Length (BL) = 4.



The first 4 data bits come from Bank 0; the next 4 from Bank 1; and the final 4 from Bank 2.

Note that a new term, READ Latency, has been introduced. READ Latency is the sum of Additive Latency and CAS Latency; RL = AL + CL. Note also that, if the READ command follows the ACTIVE command by n clock cycles, the access time from the ACTIVE operation, or the random access time, is RL + n. Clearly, the minimum value of n is 1 and occurs when the ACTIVE and READ commands are issued on successive clock cycles. Thus, the minimum value of random access time is RL + 1.

[19] Ibid., pp. 30-33

Copyright© 2022 William R. Huber

4. Other DDR2 Changes

a. Extended Mode Registers 2 and 3

The DDR2 architecture introduces two more mode registers, Extended Mode Registers 2 and 3. However, only one bit in EMR2 actually carries any meaningful information. Bit 7 specifies a faster refresh rate for operation above 85°C.

b. 4n-bit Prefetch

Footnote 8 on page 8 briefly discusses the 2n-bit prefetch used in DDR SDRAMs. Prefetch refers to the function of accessing bits from the selected columns of memory during a READ operation. Because of internal memory speed limitations, enough bits must be accessed to satisfy the first few data outputs. At DDR clock speeds, 2 bits per DQ pin are sufficient. At the faster DDR2 clock speeds, 4 bits per DQ pin are required. The "n" in 2n or 4n refers to the number of DQ pins. So a 4n-bit prefetch on a x16 DDR2 SDRAM requires that 64 bits of data be prefetched.

c. CAS Latency

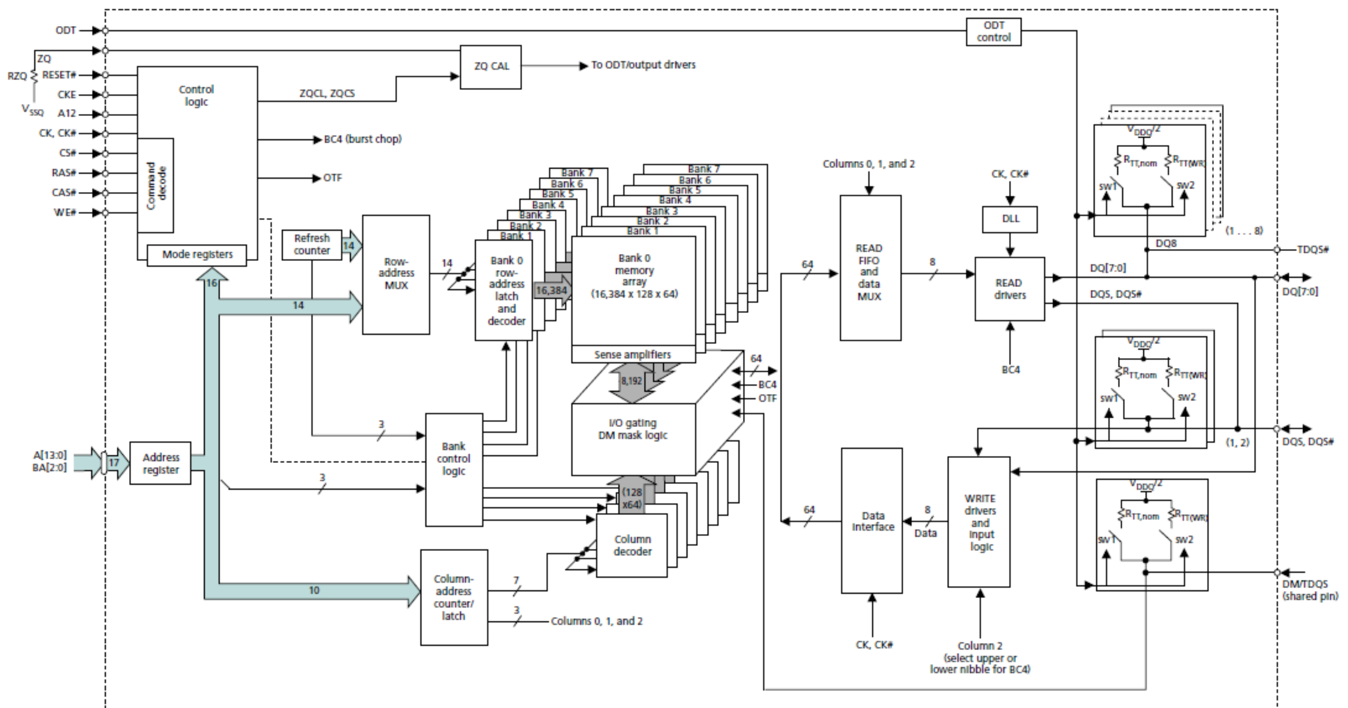
CAS Latency has the same meaning in both DDR and DDR2 SDRAMs. However, because of the higher DDR2 clock speeds, CAS Latency values are greater for DDR2 than for DDR devices. In addition, DDR2 SDRAMs do not support half-clock cycle CAS Latency values.

 Copyright© 2022 William R. Huber

D. Double Data Rate III (DDR3) SDRAMs

JEDEC, the standard-setting body for semiconductor memory devices, issued the initial version of the DDR3 SDRAM standard in June 2007.[20] At that time, Micron expected DDR3 to be economically and technically viable from 2007 to 2014, and to be the dominant architecture after 2010.[21] Let's look at a block diagram of a DDR3 device:[22]

Figure 4: 128 Meg x 8 Functional Block Diagram



[20] JEDEC JESD79-3
[21] "DDR3 Advantages," Micron Marketing, 4/8/2009
[22] Micron MT41J128M8 1Gb DDR3 SDRAM Data Sheet, Rev. J 05/10 EN, page 15

Copyright© 2022 William R. Huber

At this high level, there are a few changes visible relative to the DDR2 block diagram.

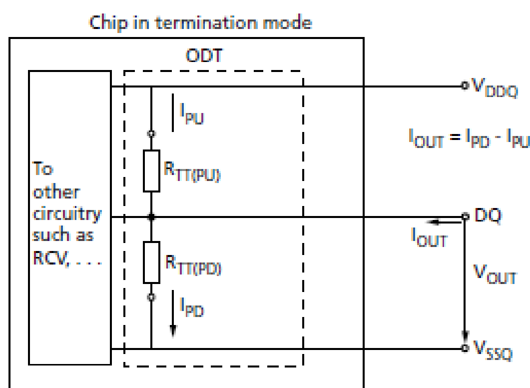
- An external RZQ resistor and associated internal ZQ CAL block are added.
- There are 8 banks as opposed to 4 banks in the DDR2 device we examined. Actually, higher density (1 Gb and larger) DDR2 devices also have 8 banks.
- 8n-bit prefetch is used instead of the 4n-bit prefetch in DDR2 devices.

1. On-Die Termination in DDR3

Starting on page 19, we discussed how the DDR2 architecture introduced on-die termination (ODT) with effective resistance values of 50, 75, 150 or infinite ohms to improve signal integrity, primarily when receiving data during WRITE operations. The DDR3 architecture expands on the ODT capability in two ways: Calibration Based on External Precision Resistor; and Dynamic ODT.

a. Calibration Based on External Precision Resistor

As with the DDR2 architecture, the output terminating circuit consists of two resistors, $R_{TT(PU)}$ tied between the DQ pin and the positive power supply voltage, V_{DDQ}; and $R_{TT(PD)}$ tied between the DQ pin and ground, V_{SSQ}. The diagram is to the left.[23]



The effective terminating resistance, $R_{TT(EFF)}$, is the parallel combination of the two resistors. $R_{TT(EFF)}$ can be 20, 30, 40, 60, 120 or infinite ohms; and is selected by bits 2, 6 and 9 of Mode Register 1 (MR1).[24] As in DDR2, the state of the ODT pin when the command is registered determines whether or not the resisters are connected to their corresponding power supplies.

In an attempt to provide more accurate values for the terminating resistors, DDR3 provides for internal resistor calibration via a precision external resistor, RZQ. When a 240Ω±1% resistor is connected from the ZQ pin to ground, the ZQ CALIBRATION[25] command is used to calibrate both the output

[23] Ibid, page 55

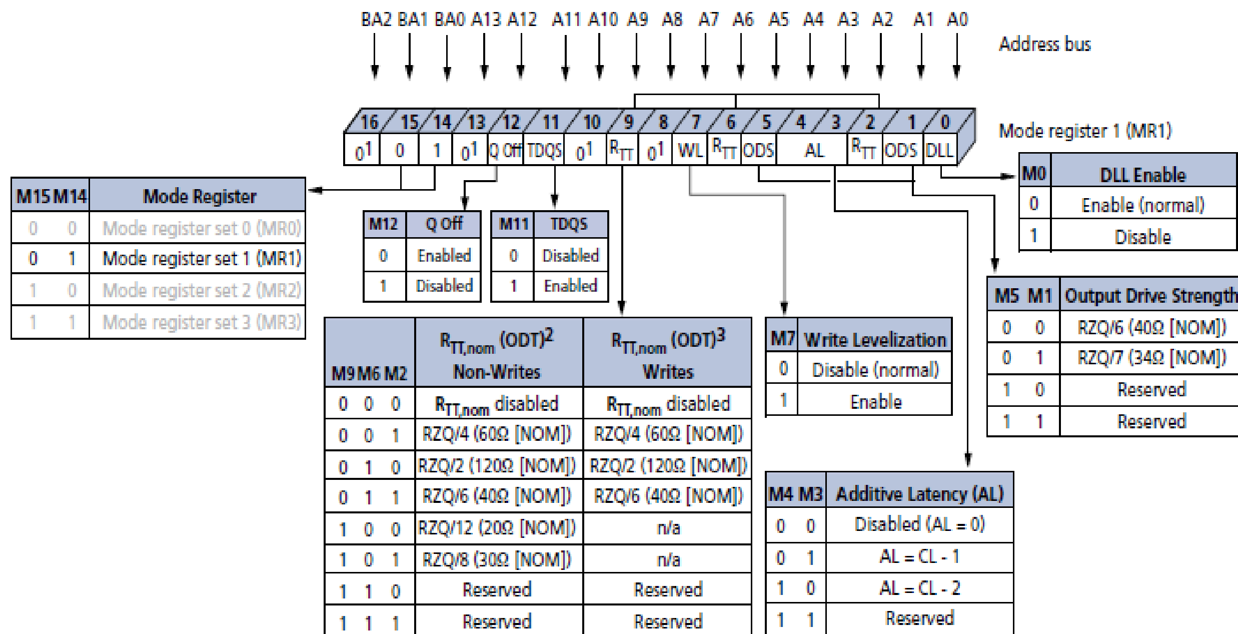
[24] The Mode Register and Extended Mode Registers of the DDR2 architecture have been replaced by Mode Registers MR0 to MR3 in the DDR3 architecture.

[25] There are two ZQ CALIBRATION commands: ZQCL for initial calibration and for calibration after exiting a SELF-REFRESH operation; and ZQCS for periodic recalibrations.

 Copyright© 2022 William R. Huber

termination resistors (used primarily for WRITE operations) and the output driver impedance (R_{ON}, used for READ operations) over process, voltage and temperature variations. R_{ON} is either 34 or 40 ohms, as selected by bits 1 and 5 of MR1, as shown below:[26]

Figure 55: Mode Register 1 (MR1) Definition



M15	M14	Mode Register
0	0	Mode register set 0 (MR0)
0	1	Mode register set 1 (MR1)
1	0	Mode register set 2 (MR2)
1	1	Mode register set 3 (MR3)

M12	Q Off
0	Enabled
1	Disabled

M11	TDQS
0	Disabled
1	Enabled

M0	DLL Enable
0	Enable (normal)
1	Disable

M5	M1	Output Drive Strength
0	0	RZQ/6 (40Ω [NOM])
0	1	RZQ/7 (34Ω [NOM])
1	0	Reserved
1	1	Reserved

M9	M6	M2	$R_{TT,nom}$ (ODT)² Non-Writes	$R_{TT,nom}$ (ODT)³ Writes
0	0	0	$R_{TT,nom}$ disabled	$R_{TT,nom}$ disabled
0	0	1	RZQ/4 (60Ω [NOM])	RZQ/4 (60Ω [NOM])
0	1	0	RZQ/2 (120Ω [NOM])	RZQ/2 (120Ω [NOM])
0	1	1	RZQ/6 (40Ω [NOM])	RZQ/6 (40Ω [NOM])
1	0	0	RZQ/12 (20Ω [NOM])	n/a
1	0	1	RZQ/8 (30Ω [NOM])	n/a
1	1	0	Reserved	Reserved
1	1	1	Reserved	Reserved

M7	Write Levelization
0	Disable (normal)
1	Enable

M4	M3	Additive Latency (AL)
0	0	Disabled (AL = 0)
0	1	AL = CL - 1
1	0	AL = CL - 2
1	1	Reserved

b. Dynamic ODT

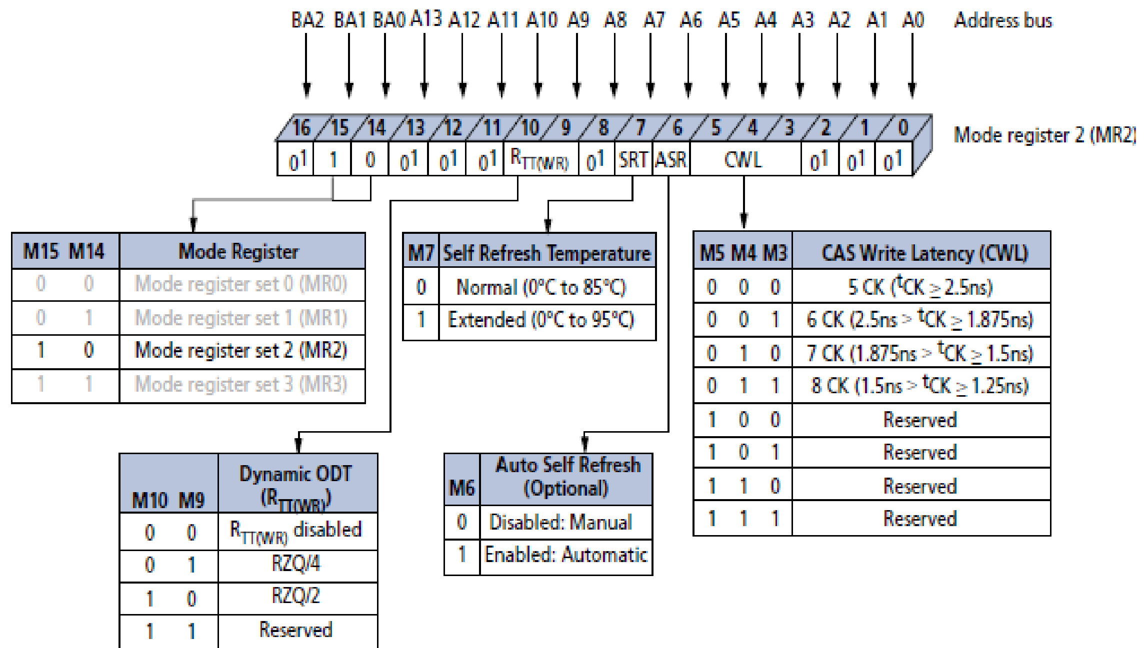
Some memory applications require different values of terminating resistances during WRITE operations as opposed to other operations (or idle times). In DDR3, these different values can be achieved without having to reload the bits in MR1 by using dynamic ODT. With dynamic ODT enabled, the termination resistance changes from its nominal value ($R_{TT,nom}$) to its dynamic value ($R_{TT(WR)}$) when beginning a WRITE burst; and subsequently switches back to $R_{TT,nom}$ when the WRITE burst is complete. For example: $R_{TT,nom}$ could be set to 40 ohms (MR1 bits 9-6-2 storing 0-1-1) and $R_{TT(WR)}$ could be set to 120 ohms (MR2 bits 10-9 storing 1-0). Then the termination resistance would start at 40 ohms; switch to 120 ohms during the WRITE burst; then return to 40 ohms after the WRITE burst is complete, all without the need for a MODE REGISTER SET command.

[26] Micron MT41J128M8 1Gb DDR3 SDRAM Data Sheet, Rev. J 05/10 EN, page 135

 Copyright© 2022 William R. Huber

The organization of Mode Register 2 (MR2) is shown below.[27] Dynamic ODT is enabled if either bit 9 or bit 10 (but not both) is a 1.

Figure 57: Mode Register 2 (MR2) Definition



2. CAS Write Latency

We discussed Additive Latency (AL) in DDR2 memories starting on page 22. In that context, Read Latency, the time in clock cycles from the READ command to valid data is equal to Additive Latency + CAS Latency; or, RL = AL + CL.

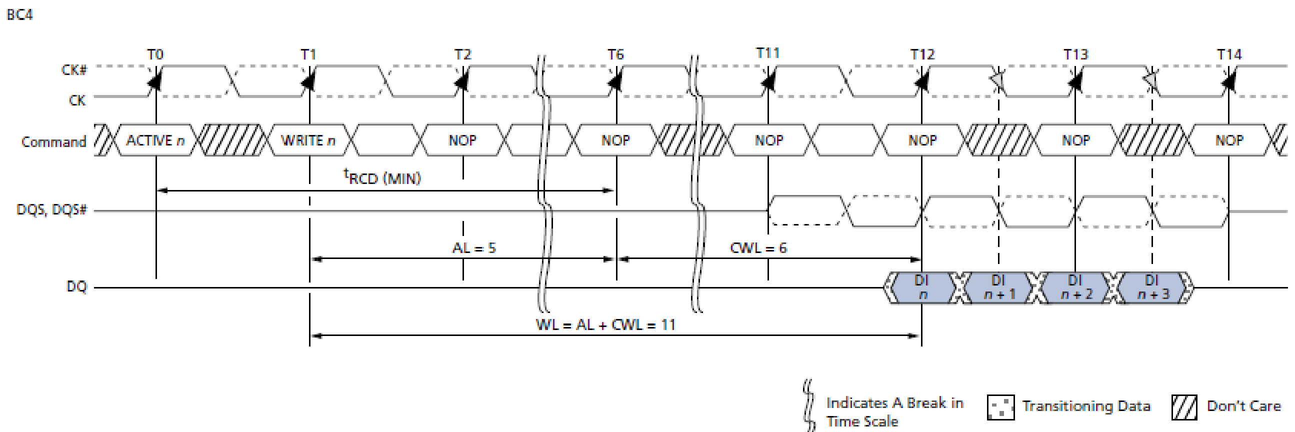
DDR3 introduces a new term, CAS Write Latency (CWL). CWL is the time from the expiration of Additive Latency until the first input data during a WRITE operation can be accepted by the memory. As shown above, information defining CWL is stored in bits 5, 4 and 3 of Mode Register 2. A timing diagram illustrating CAS Write Latency is shown below.[28]

---

[27] Ibid., page 139
[28] Loc. Cit.

 Copyright© 2022 William R. Huber

Figure 58: CAS Write Latency



For this diagram:

- ACTIVE command is registered at Time = T0;
- posted WRITE command is registered at Time = T1;
- Additive Latency (AL) = 5;
- AL expires at Time = T6;
- CAS Write Latency (CWL) = 6;
- first input data is accepted at Time = T12; and
- overall Write Latency (WL) = AL + CWL = 11.

3. 8n-bit Prefetch

As we discussed in conjunction with DDR and DDR2 memories, prefetch refers to the function of accessing bits from the selected columns of memory during a READ operation. Because of internal memory speed limitations, enough bits must be accessed to satisfy the first few data outputs. At DDR clock speeds, 2 bits per DQ pin are sufficient. At the faster DDR2 clock speeds, 4 bits per DQ pin are required. The "n" in 2n or 4n refers to the number of DQ pins. With even faster DDR3 devices, 8 bits per DQ pin are required. So the 8n-bit prefetch on a x16 DDR3 SDRAM requires that 128 bits of data be prefetched. Thus the internal data bus on DDR3 memories is extremely wide.

 Copyright© 2022 William R. Huber

E. Double Data Rate IV (DDR4) SDRAMs[29]

JEDEC issued the initial version of the DDR4 SDRAM standard in September 2012. It was most-recently updated in June 2017. Main memory (DRAM) is still all about capacity and data transfer rate, and DDR4 improves over DDR3 in both areas. The largest DDR4 memories store 16Gb of information (vs. 8Gb for DDR3); and the fastest speed grade for DDR4 is 3200MHz (vs. 2133MHz for DDR3).

But DDR4 also adds several new features, many of them aimed at reducing power consumption:

- The external power supply is reduced from 1.5V to 1.2V. This step, which takes advantage of the smaller feature sizes of advanced processing, significantly reduces power consumption for an equivalent memory device.
- DIMM interface (see pp. 38-41) is 284-pins, vs. 240-pins for DDR3. The increased pins support more addresses and allow a 1:1 ratio of signal-to-ground pins (DDR3 was 2:1).
- A dedicated Activate input, ACT_n, is added in DDR4; and other control inputs are multiplexed to provide row address inputs (RAS_n/A16, CAS_n/A15 and WE_n/A14). The net effect is to save two input pins/balls.
- An added level of addressing hierarchy, Bank Groups, allows DDR4 memories to have separate activation, read, write or refresh operations underway simultaneously in each unique bank group.
- Data output is pseudo-open drain (POD) in DDR4, vs. series-stub terminated logic (SSTL) in DDR3. The POD structure reduces current in the output drivers, especially when driving a high to the data bus.[30]
- VPP (used to provide high word-line voltages to effectively activate the cell access transistors) is supplied externally in DDR4. This saves power as compared to the internally-generated (charge pump) VPP voltage in DDR3.
- Parity detection of commands and addresses is supported in DDR4; no such capability exists in DDR3. Detection of a parity error results in blocking the erroneous command instead of executing it.

---

[29] See JEDEC JESD79-4B (June 2017) for details of the DDR4 SDRAM specification
[30] Wang, David; "Why migrate to DDR4?"; March 12, 2013

 Copyright© 2022 William R. Huber

- Internal cyclic redundancy check (CRC) is computed and validated to verify write data in DDR4. No such capability exists in DDR3.
- In addition to writing to them, the internal mode registers in DDR4 can be read externally. In DDR3 (and previous versions), the mode registers cannot be read externally.
- DDR4 supports low-power programmable self-refresh, which varies the refresh rate depending on die temperature and device activity. Longer refresh intervals are acceptable if die temperature is low, thus saving power.
- Data Bus Inversion (DBI) allows either true or inverted data to be stored. In a large system (think server farms), this feature can reduce power consumption and improve data signal integrity.

The power-saving aspects of the above features, along with internal architecture changes, are expected to reduce DDR4 power consumption by 30% to 40% vs. an equivalent DDR3 device. Alternatively, at the same power, it is expected that a DDR4 device can operate 1.7X faster than the equivalent DDR3 memory.[31]

F. Double Data Rate V (DDR5) SDRAMs

By mid-2020, the once-blazing speed of DDR4 was starting to be too slow for artificial intelligence and machine learning applications. Enter DDR5, which starts at 3200MHz (the top of the DDR4 range) and extend up to 6400MHz.

Power consumption is always a problem with high-speed operation, and DDR5 addresses that by slightly reducing power supply voltage from 1.2V to 1.1V.

Beyond increased speed and lower power consumption, the most important added feature of DDR5 is called "Same Bank Refresh." In DDR4 and previous generations, whenever any one of the banks (16 in DDR4) required refreshing, all banks had to be refreshed simultaneously. Therefore, reading from or writing to the memory had to wait for the refresh operation to complete.

DDR5 chips have 32 banks. But more important, with Same Bank Refresh, every bank refreshes independently. So, while one bank is refreshing, the other 31 banks are available to the system for reading or writing. Thus, refresh overhead is reduced to almost nothing.

[31] Rastgar, Fred and Tom Rossi; "Addressing the challenges of transition to DDR4"; January 22, 2013

 Copyright© 2022 William R. Huber

Another new feature on DDR5 chips is on-die ECC (Error Correction Code). With ECC, the memory chip detects and corrects single bit errors.

The JEDEC standard for DDR5 is titled JESD-79.5[32] and costs $369. A free data sheet from Micron is available and occupies 529 pages![33]

As of April 2022, Micron, SK Hynix, and Samsung were shipping DDR5 SDRAMs.

[32] https://www.jedec.org/standards-documents/docs/jesd79-5a
[33] https://media-www.micron.com/-/media/client/global/documents/products/data-sheet/dram/ddr5/ddr5_sdram_core.pdf?rev=f76eb9631b674e66a2026c324a95cb67

 Copyright© 2022 William R. Huber

G. Comparison of SDRAMs

The following table compares many characteristics of several of the various memory devices discussed above.[34,35]

Feature/Option	SDRAM	DDR	DDR2	DDR3	DDR4	DDR5[36,37]
Data Transfer Rate (Mega-Transfers per second, MT/s)	66, 100, 133, 143	200, 266, 333, 400	400, 533, 667, 800, 1000, 1066, 1200	800, 1066, 1333,1600, 1866, 2133	1600, 1866, 2133, 2400, 2666, 3200	3200 to 6400 (6800 to 8400 future)
Power Supply (core and I/O)	3.3V	2.5V	1.8V	1.5V	1.2V (and 2.5V for VPP)	1.1V (and 1.8V for VPP)
Densities	64Mb-512Mb	64Mb-1Gb	256Mb-2Gb	512Mb-8Gb	2Gb-16Gb	8Gb-64Gb
Internal Banks	4	4	4 or 8	8	16	32
Bank Groups	N/A	N/A	N/A	N/A	2 or 4	4 or 8
Prefetch	1n	2n	4n	8n	8n	16n
CAS Latency (CL)	2, 3	2, 2.5, 3	3, 4, 5, 6	6-14[38]	11-22[34]	22-66
Additive Latency (AL)	No	No	0, 1, 2, 3, 4	0, CL - 1 or CL - 2	CAL[39] = 3, 4, 5, 6 or 8	N/A
READ Latency (RL)	CL	CL	AL + CL	AL + CL	AL + CL	CL
CAS WRITE Latency (CWL)	None	None	None	5, 6, 7, 8	9-12, 14-16, 18, 20[34]	CL - 2
WRITE Latency (WL)	Fixed	Fixed	READ Latency - 1	AL + CWL	AL + CWL	CWL
DQ Width	x4, x8, x16	x4, x8, x16	x4, x8, x16	x4, x8, x16	x4, x8, x16	x4, x8, x16
On-Die Termination (ODT) (ohms)	None	None	50, 75, 150, ∞	Fixed (20, 30, 40, 60, 120, ∞); Dynamic (60 or 120)	34, 40, 48, 60, 80, 120, 240 (disables during READ bursts)	34, 40, 48, 60, 80, 120, 240 (disables during Self Refresh)
Burst Length (BL)	1, 2, 4, 8, full page	2, 4, 8	4, 8	4C (chop; last 4 bits of 8 burst masked), 8	4C (chop; last 4 bits of 8 burst masked), 8	8, 16, 32

---

[34] Partially from Micron TN-47-02, "DDR2 Offers New Functions/Functionality"
[35] Partially from "DDR4 – Advantages of Migrating from DDR3" https://www.micron.com/products/dram/ddr3-to-ddr4
[36] Partially from Proposed DDR5 Full Spec, JESD79.5
[37] From Micron DDR5 Product Core Data Sheet https://media-www.micron.com/-/media/client/global/documents/products/data-sheet/dram/ddr5/ddr5_sdram_core.pdf?rev=f76eb9631b674e66a2026c324a95cb67
[38] Depending on speed grade
[39] CAL is Command/Address Latency, the delay from CK rising to Command or Address detection

Copyright© 2022 William R. Huber

G. What Comes Next?

In 2011, JEDEC published the Wide I/O 2 standard; it stacks multiple memory dice but does that directly on top of the CPU and in the same package. This memory layout provides higher bandwidth and better power performance than DDR4 SDRAM, and allows a wide interface with short signal lengths. It primarily aims to replace various mobile DDRX SDRAM standards used in high-performance embedded and mobile devices, such as smartphones.

Hynix proposed a similar High Bandwidth Memory (HBM), which was published as JEDEC JESD235. Both Wide I/O 2 and HBM use a very wide parallel memory interface, up to 512 bits wide for Wide I/O 2 (compared to 64 bits for DDR4), running at a lower frequency than DDR4. Wide I/O 2 is targeted at high-performance compact devices such as smartphones, where it will be integrated into the processor or system on a chip (SoC) packages.

Micron Technology's Hybrid Memory Cube (HMC) stacked memory uses a serial interface. Many other computer buses have migrated towards replacing parallel buses with serial buses, for example by the evolution of Serial ATA replacing Parallel ATA, PCI Express replacing PCI, and serial ports replacing parallel ports. In general, serial buses are easier to scale up and have fewer wires/traces, making circuit boards using them easier to design.

HBM is targeted at graphics memory and general computing, while HMC targets high-end servers and enterprise applications.[40]

H. RLDRAMs

The RL in RLDRAM stands for Reduced Latency, and these devices are optimized for fast random-access applications where read and write cycles are interspersed. The RLDRAM family was co-developed by Micron and Infineon Technologies; but with the spin-out of Qimonda by Infineon and the subsequent bankruptcy filing of Qimonda, Micron is now the primary vendor of the RLDRAM architecture. In January 2011, Renasas announced a low-latency DRAM[41] similar to the RLDRAM2 memory described in the table below.

JEDEC has not standardized the RLDRAM architecture, so the features of RLDRAMs are more variable than those of DDR SDRAMs. For example, some

---

[40] https://en.wikipedia.org/wiki/DDR4_SDRAM
[41] http://www.eetimes.com/electronics-products/electronic-product-reviews/memory-products/4212399/Renesas-releases-DRAM-for-large-capacity-networks?cid=NL_Memory

 Copyright© 2022 William R. Huber

RLDRAMs have shared input/output pins, as do all DDR SDRAMs. But other RLDRAMs have separate input and output pins.

RLDRAMs can accept addresses in two ways:

(1) as multiplexed row and column addresses (all addresses appear on the same set of pins but are separated in time), the same as DDR SDRAMs; or

(2) all addresses at once on dedicated address pins, similar to high speed static RAMs.

The following table compares features of DDR3 SDRAMs with the latest two versions of Micron RLDRAMs.[42]

Feature/Parameter	DDR3 SDRAM	RLDRAM 2	RLDRAM 3
Data Transfer Rate	800-2133 MHz	350-1066 MHz	400-2133 MHz
Power Supply	1.5V; 1.35V Core & IO	1.8V Core; 1.5-1.8V IO	1.35V Core; 1.2V IO
Storage Capacity	512Mb to 8Gb	288 and 576Mb	576Mb and 1Gb
Internal Banks	8	8	16
t_{RC} (Activate-to- Activate Cycle Time)	46-52 ns	15-20 ns	<10 ns
Data Width	x4, x8, x16	x9, x18, x36	x18, x36
On-Die Termination (ODT)	Programmable	Programmable	Programmable
Burst Length (BL)	4C[43], 8	2, 4, 8	2, 4, 8
External Timing Signals	CK, CK#	CK, CK#; DK, DK# (for data-in)	CK, CK#; DK, DK# (for data-in)
Internal Fault Detection	None	JTAG Boundary Scan	JTAG Boundary Scan

[42] Partially from "RLDRAM Forges Ahead," www.micron.com/products/rldram3.html and "RLDRAM--Built for Networking," www.micron.com/products/dram/rldram.html

[43] For 4-bit burst, 8 bits are accessed, but the outputs of the last 4 bits are masked

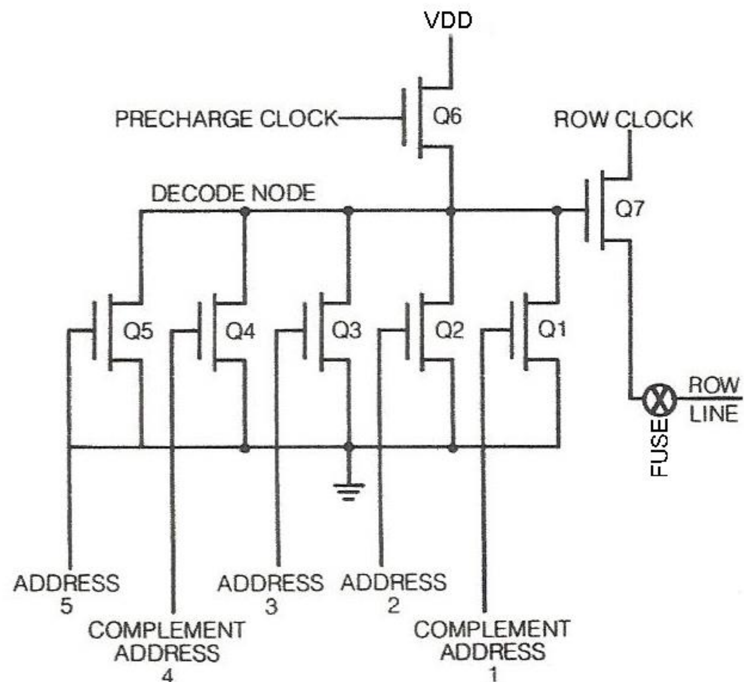
 Copyright© 2022 William R. Huber

I. Inside and Outside the Memory Chip

1. Redundancy

Memory chips are physically large and internally complex. A 1Gb memory chip contains the equivalent of about 2.3 x 10^9 transistors. Because of this size and complexity, manufacturing defects can severely limit the number of good chips per wafer. Fortunately, the repetitive row by column nature of memory devices provides the opportunity to use redundant rows and columns of memory cells to replace defective rows and columns. Since about 1985, virtually every DRAM has used redundancy to improve manufacturing yields. To understand how redundancy is implemented, consider this simplified diagram of a "normal" row decoder.[44]

Row decoders receive binary codes from address buffer circuits via the internal address lines and complement address lines. These codes, made up of addresses and complement addresses in unique combinations, allow the proper row to be selected during each memory operation. During memory operation, an address and a complement address are binary opposites. If, for example, an address is a 0 or a low voltage, its complement would be a 1 or a high voltage. The complete memory contains many decoders, differing only in the addresses and complement addresses that control them.



In the standby condition, when the memory is simply storing information, the address and complement address lines are at 0 volts. Therefore, transistors Q1 through Q5 are not conducting. The decoder circuit must be "loaded" or precharged during this time, when the circuit is not in use. The precharge clock turns on and the decode node is charged to VDD volts through transistor Q6.

[44] Huber, W.R., "The 64K RAM: a fault-tolerant semiconductor memory design," ***Bell Laboratories Record***, July/August 1979, pp. 199-204

 Copyright© 2022 William R. Huber

The voltage on the decode node turns on Q7, thereby connecting the row line to the row clock. The row clock—which sends out pulses to activate the transistors of the memory cells—is at 0 volts when the memory is in standby. With the row line at 0 volts, all memory cells along the line are storing information but cannot be accessed.

A series of events takes place when a memory is accessed. First, the precharge clock switches to 0 volts and turns off Q6. The decode node voltage is unaffected and remains at VDD volts.

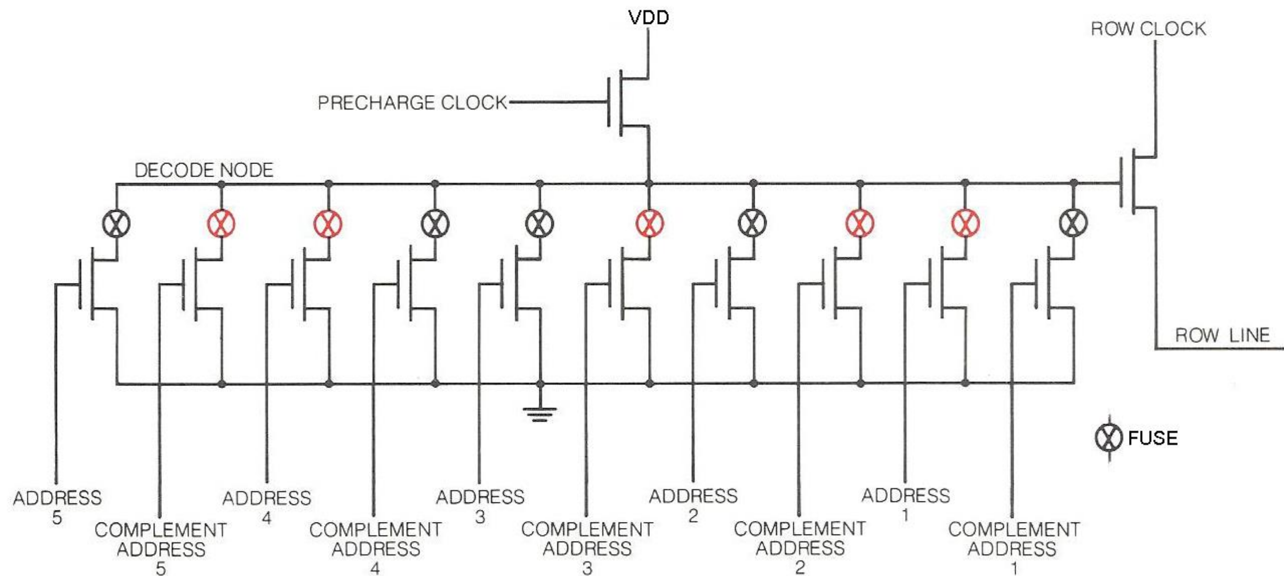
Next, the address and complement address signals are sent to the decoder on the address lines and complement address lines. Only one unique combination of external address voltages—in this case represented by 0-1-0-0-1—will keep all the address lines and complement address lines in the figure at 0 volts. For this combination, transistors Q1 through Q5 remain off and the decode node stays at VDD volts. This is the "selected" state. For all other address signals, one or more of the address and complement address lines will go to a high voltage. This causes one or more of transistors Q1 through Q5 to discharge the decode node to 0 volts—a condition called the "deselected" state. During a single memory access operation the address voltages remain fixed and only one decoder circuit is selected; all others are deselected.

Finally, when the row clock switches on, transistor Q7 of the selected decoder circuit connects the row clock voltage to the row line. This voltage then activates, or permits access to, the memory cells along the row.

If a row or a part of a row of memory cells proves defective during testing, the polysilicon programmable fuse in series with that row line can be cut open by a laser beam. The opened fuse blocks the row clock voltage from the row line.

Once the nonworking row is eliminated, a spare row has to be programmed to take its place. A simplified spare row decoder is shown in the following figure.

 Copyright© 2022 William R. Huber



This spare row decoder works the same way as the normal row decoder except that, until it's programmed, the spare decoder is controlled by every binary address *and* its complement address.

Because either an address or its complement goes to a high voltage during every memory operation, the spare decoder is always deselected. However, when the proper laser-programmable fuses are opened between the decode node and the transistors, the spare decoder can be made to select for any desired combination of addresses and complement addresses. If, for example, the red links are opened, the spare decoder will behave exactly like the normal decoder shown on p. 31—it will be selected by the binary code 0-1-0-0-1. Thus, the spare row of memory cells is accessed instead of the disconnected normal row.

Similar circuitry allows defective columns to be eliminated and spare columns to be substituted in their places.

 Copyright© 2022 William R. Huber

2. Memory Packaging and Speed Identification

Memories in single chip packages are rarely sold to consumers. Instead, small circuit boards called Dual-Inline Memory Modules (DIMMs) containing 8, 9, 16 or 18 memory chips in small packages, possibly along with other devices, are readily available. Here is what a DDR DIMM looks like; it is about 5.25" long and 1.2" high.

Note that the DIMM pictured above carries the identification "DDR 512/400." This means that it is a 512MB DIMM composed of memory chips capable of running at a data bus and strobe frequency of 400 MHz. Thus, it is a PC3200 DIMM.

For each DRAM generation, there are certain bus clock frequencies that are standardized. There is also a DIMM nomenclature for each of these speeds for each type. DIMMs for different generations are not compatible with each other, either physically or electrically. Basic DIMM characteristics are tabulated on the next page.[45]

The maximum storage capacity on a single DIMM is 128GB for DDR3 and 512GB for DDR4.[46]

Appendix A provides more detail on speed information that is sometimes included on DIMMs.

---

[45] Adapted from Wikipedia, http://en.wikipedia.org/wiki/DIMM
[46] Wang, op cit.

 Copyright© 2022 William R. Huber

SDR SDRAM DIMMs (100, 144 or 168 pins)				
Chip	**Module**	**Clock**	**Transfer rate**	**Voltage**
SDR-66	PC-66	66 MHz	66 MT/s	3.3 V
SDR-100	PC-100	100 MHz	100 MT/s	3.3 V
SDR-133	PC-133	133 MHz	133 MT/s	3.3 V

DDR SDRAM (DDR1) DIMMs (172, 184 or 200 pins)					
Chip	**Module**	**Memory Clock**	**I/O Bus Clock**	**Transfer rate**	**Voltage**
DDR-200	PC-1600	100 MHz	100 MHz	200 MT/s	2.5 V
DDR-266	PC-2100	133 MHz	133 MHz	266 MT/s	2.5 V
DDR-333	PC-2700	166 MHz	166 MHz	333 MT/s	2.5 V
DDR-400	PC-3200	200 MHz	200 MHz	400 MT/s	2.5 V
DDR2 SDRAM DIMMs (200, 214, 240 or 244 pins)					
Chip	**Module**	**Memory Clock**	**I/O Bus Clock**	**Transfer rate**	**Voltage**
DDR2-400	PC2-3200	200 MHz	200 MHz	400 MT/s	1.8 V
DDR2-533	PC2-4200	266 MHz	266 MHz	533 MT/s	1.8 V
DDR2-667	PC2-5300	333 MHz	333 MHz	667 MT/s	1.8 V
DDR2-800	PC2-6400	400 MHz	400 MHz	800 MT/s	1.8 V
DDR2-1066	PC2-8500	533 MHz	533 MHz	1066 MT/s	1.8 V
DDR3 SDRAM DIMMs (204 or 240 pins)					
Chip	**Module**	**Memory Clock**	**I/O Bus Clock**	**Transfer rate**	**Voltage**
DDR3-800	PC3-6400	400 MHz	400 MHz	800 MT/s	1.5 V
DDR3-1066	PC3-8500	533 MHz	533 MHz	1066 MT/s	1.5 V
DDR3-1333	PC3-10600	667 MHz	667 MHz	1333 MT/s	1.5 V
DDR3-1600	PC3-12800	800 MHz	800 MHz	1600 MT/s	1.5 V
DDR3-1866	PC3-14900	933 MHz	933 MHz	1866 MT/s	1.5 V
DDR3-2133	PC3-17000	1066 MHz	1066 MHz	2133 MT/s	1.5 V
DDR3-2400	PC3-19200	1200 MHz	1200 MHz	2400 MT/s	1.5 V
DDR4 SDRAM DIMMs (260, 284 or 288 pins)					
Chip	**Module**	**Memory Clock**	**I/O Bus Clock**	**Transfer rate**	**Voltage**
DDR4-1600	PC4-12800	800 MHz	800 MHz	1600 MT/s	1.2 V
DDR4-1866	PC4-14900	933 MHz	933 MHz	1866 MT/s	1.2 V
DDR4-2133	PC4-17000	1066 MHz	1066 MHz	2133 MT/s	1.2 V
DDR4-2400	PC4-19200	1200 MHz	1200 MHz	2400 MT/s	1.2 V
DDR4-2666	PC4-21300	1333 MHz	1333 MHz	2666 MT/s	1.2 V
DDR4-3200	PC4-25600	1600 MHz	1600 MHz	3200 MT/s	1.2 V

Copyright© 2022 William R. Huber

Appendix A
DIMM Timing Information

Some DIMMs carry additional speed information in the form of four numbers separated by dashes, such as 7-6-5-15. What do those cryptic numbers mean?

- The first number (7 in the example) denotes the CAS Latency of the memory devices on the DIMM. As we have discussed, CAS Latency defines the time from the registration of the READ command to the appearance of valid data on the Data Input/Output (DQ) pins and is measured in clock cycles.

- The second number (6 in the example) denotes t_{RCD}, the minimum delay in clock cycles between the ACTIVE command and a READ or WRITE command for an SDRAM or DDR device.

  Because of Additive Latency, the interpretation is a bit more complicated for a DDR2 or DDR3 device. As we have discussed, in these devices the ACTIVE command and READ (or WRITE) command can be issued on successive clock cycles, but the READ (WRITE) command is not acted on by the device until the expiration of the Additive Latency (AL) time. Thus, the effective time between the ACTIVE command and the READ (WRITE) command is AL + 1 clock cycles (assuming the ACTIVE command and the READ (WRITE) command are issued on successive clock cycles). So the second number in the speed information is equal to AL + 1; or AL = (second number) -1. Of course, AL can be set to a value larger than (second number) -1 at a sacrifice in performance.

- The third number (5 in the example) denotes the memory timing parameter t_{RP}, in clock cycles. On page 8 we briefly discussed Auto-Precharge, an option for the READ (WRITE) command that automatically returns the memory to its idle state. If Auto-Precharge is not selected, then a separate PRECHARGE command must be issued after the READ (WRITE) command to return the memory to its idle state (and thus be ready for another ACTIVE

 Copyright© 2022 William R. Huber

command). t_{RP} is the minimum time between the registration of the PRECHARGE command and the registration of the next ACTIVE command.

- The fourth number (15 in the example) denotes the memory timing parameter t_{RAS}, in clock cycles. t_{RAS} is the time between the registration of the ACTIVE command and the registration of the next PRECHARGE command. Note that the memory cycle time (time in clock cycles between ACTIVE commands) is given by the sum of the third and fourth numbers in the speed information. For the example, the memory cycle time is 20 clock cycles.

Now we can calculate the various timing parameters discussed above from the DIMM speed information. To achieve a generic result, assume the speed information is A-B-C-D. Most DDR2 and DDR3 entries assume that ACTIVE and READ (WRITE) commands are registered on successive clock cycles.

Parameter	Symbol	DDR	DDR2	DDR3
CAS Latency	**CL**	A	A	A
ACTIVE to READ (WRITE)	t_{RCD}	B	AL + 1	A or A - 1 (= AL + 1)
Additive Latency	**AL**	---	B - 1	A - 1 or A - 2[47]
Read Latency	**RL**	A	A + B - 1	2A - 1 or 2A - 2 (= CL + AL)
Write Latency	**WL**	---	A + B - 2	A - 1 + CWL[48] or A - 2 + CWL (= AL + CWL)
Memory Cycle Time	t_{RC}	C + D	C + D	C + D

[47] In the DDR3 architecture, AL is limited to Disabled, CL - 1 or CL - 2
[48] CWL is the CAS Write Latency, and is discussed on pp. 28-29

 Copyright© 2022 William R. Huber

Cited References

1. Micron MT48LC128M4A2 512Mb SDRAM Data Sheet, Rev. M 6/10 EN
2. Micron MT46V128M4 512Mb DDR SDRAM Data Sheet, Rev. B 2/09 EN
3. Foss, R.C. et al; “Delay Locked Loop Implementation in a Synchronous Dynamic Access Memory;” U. S. Patent 6,205,083; Issued March 20, 2001
4. “DDR3 Advantages,” Micron Marketing, 4/8/2009
5. Micron MT47H64M8 512Mb DDR2 SDRAM Data Sheet, Rev. Q 10/10 EN
6. Micron TN-47-02 “DDR2 Offers New Features/Functionality”
7. “DDR3 SDRAM Standard,” JEDEC JESD79-3, June 2007
8. Micron MT41J128M8 1Gb DDR3 SDRAM Data Sheet, Rev. J 05/10 EN
9. “DDR4 SDRAM Standard,” JEDEC JESD79-4B, June 2017
10. Wang, David; “Why migrate to DDR4?”; March 12, 2013
11. Rastgar, Fred and Tom Rossi; “Addressing the challenges of transition to DDR4”; January 22, 2013
12. Patrizio, Andy; "DDR5 memory promises a significant speed boost;" NETWORKWORLD; April 12, 2020; https://www.networkworld.com/article/3537328/ddr5-memory-promises-a-significant-speed-boost.html
13. https://en.wikipedia.org/wiki/DDR4_SDRAM
14. http://www.eetimes.com/electronics-products/electronic-product-reviews/memory-products/4212399/Renesas-releases-DRAM-for-large-capacity-networks?cid=NL_Memory
15. “RLDRAM Forges Ahead,” www.micron.com/products/rldram3.html
16. “RLDRAM—Built for Networking,” www.micron.com/products/dram/rldram.html
17. Huber, W.R., “The 64K RAM: a fault-tolerant semiconductor memory design,” ***Bell Laboratories Record***, July/August 1979, pp. 199-204
18. Wikipedia, http://en.wikipedia.org/wiki/DIMM
19. www.eetimes.com/electronics-news/4211854/Samsung-trials-DDR4-DRAM-module?cid=NL_EETimesDaily
20. “DDR4 – Advantages of Migrating from DDR3” https://www.micron.com/products/dram/ddr3-to-ddr4
21. Micron DDR5 Product Core Data Sheet https://media-www.micron.com/-/media/client/global/documents/products/data-sheet/dram/ddr5/ddr5_sdram_core.pdf?rev=f76eb9631b674e66a2026c324a95cb67

 Copyright© 2022 William R. Huber

Other References

1. Prince, B., ***Semiconductor Memories***, Second Edition, ISBN 0 471 92465 2, 1983, John Wiley & Sons, West Sussex, England
2. Prince, B., ***High Performance Memories***, ISBN 0 471 95646 5, 1996, John Wiley & Sons, West Sussex, England
3. Przybylski, S.A., ***New DRAM Technologies***, Second Edition, ISBN 1-885330-06-5, 1996, MicroDesign Resources, Sebastopol, CA
4. Sharma, A.K., ***Semiconductor Memories—Technology, Testing and Reliability***, 1996, IEEE Press, New York, NY
5. Prince, B., ***Emerging Memories—Technologies and Trends***, 2002, Kluwer Academic Publishers, Norwell, MA
6. Keeth, B., R. J. Baker, B. Johnson and F. Lin, ***DRAM Circuit Design—Fundamental and High Speed Topics***, ISBN 978-0-470-18475-2, 2008, IEEE Press, New York, NY

Copyright© 2022 William R. Huber